

UNIVERSIDADE FEDERAL DE VIÇOSA

Análise espectral e modelagem preditiva para cultura da soja

Inglith de Souza Lage Alves
Doctor Scientiae

VIÇOSA - MINAS GERAIS
2025

INGLITH DE SOUZA LAGE ALVES

Análise espectral e modelagem preditiva para cultura da soja

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

Orientadora: Flora Maria de Melo Villar

Coorientadores: Felipe Lopes da Silva
Nerilson Terra Santos

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

A474a
2025

Alves, Inglith de Souza Lage, 1996-
Análise espectral e modelagem preditiva para cultura da
soja / Inglith de Souza Lage Alves. – Viçosa, MG, 2025.
1 tese eletrônica (96 f.): il. (algumas color.).

Orientador: Flora Maria de Melo Villar.
Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Engenharia Agrícola, 2025.
Inclui bibliografia.
DOI: <https://doi.org/10.47328/ufvbbt.2026.189>
Modo de acesso: World Wide Web.

1. Soja - Cultivo. 2. Soja - Melhoramento genético.
3. Análise espectral. 4. Aprendizado do computador.
5. Sensoriamento remoto. 6. Fenótipo. I. Villar, Flora Maria de
Melo, 1986-. II. Universidade Federal de Viçosa. Departamento
de Engenharia Agrícola. Programa de Pós-Graduação em
Engenharia Agrícola. III. Título.

CDD 22. ed. 633.343

INGLITH DE SOUZA LAGE ALVES

Análise espectral e modelagem preditiva para cultura da soja

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

APROVADA: 4 de novembro de 2025.

Assentimento:

Inglith de Souza Lage Alves
Autora

Flora Maria de Melo Villar
Orientadora

Essa tese foi assinada digitalmente pela autora em 07/05/2026 às 20:03:46 e pela orientadora em 29/09/2026 às 10:24:25. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **Z4Q4.19XL.ZXP5** e clique no botão 'Validar documento'.

Aos meus pais, avós e irmãos.

AGRADECIMENTOS

A Deus, fonte inesgotável de sabedoria, força e esperança, pela presença constante que me sustentou nos momentos de fraqueza e me guiou com fé até a conclusão deste trabalho.

À minha família, pelo amor incondicional, apoio, paciência e incentivo durante todos os anos de caminhada acadêmica.

À Universidade Federal de Viçosa, em especial ao Departamento de Engenharia Agrícola (DEA), pela oportunidade de realizar minha formação acadêmica em um ambiente de excelência científica.

À professora Flora Maria de Melo Villar, minha orientadora, pela orientação segura, pelas contribuições técnicas e pela confiança depositada ao longo de todas as etapas deste trabalho.

Aos professores Felipe Lopes da Silva e Nerilson Terra Santos, coorientadores, pela disponibilidade, suporte científico e pela atenção durante o desenvolvimento desta pesquisa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pelo apoio financeiro na forma de bolsa de estudos, fundamental para a realização deste trabalho.

Aos técnicos, colaboradores e colegas do DEA, pelo auxílio durante a condução dos experimentos e por tornarem essa jornada mais leve.

Aos amigos que fiz ao longo de toda a pós-graduação, pelo companheirismo, trocas de conhecimento e ajuda recebida ao longo dessa caminhada.

A todos aqueles que, de forma direta ou indireta, contribuíram para o desenvolvimento deste trabalho, deixo o meu sincero reconhecimento e profundo agradecimento.

Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

“Consagre ao Senhor tudo o que você faz, e os seus planos serão bem-sucedidos”.
(Provérbios 16:3)

RESUMO

ALVES, Inglith de Souza Lage, D.Sc., Universidade Federal de Viçosa, novembro de 2025. **Análise espectral e modelagem preditiva para cultura da soja.** Orientadora: Flora Maria de Melo Villar. Coorientadores: Felipe Lopes da Silva e Nerilson Terra Santos.

A utilização de técnicas de sensoriamento remoto e aprendizado de máquinas tem se mostrado fundamental para a compreensão e predição do desempenho agrônômico da soja, especialmente em um contexto de crescente demanda por métodos não destrutivos e de alta precisão. Neste trabalho, investigou-se a aplicabilidade de índices espectrais derivados de imagens multiespectrais adquiridas por sensores embarcados em veículos aéreos não tripulados (VANTs) para a fenotipagem remota, caracterização espectral e predição da produtividade da cultura da soja. Foram inicialmente analisadas as correlações entre nove índices de vegetação (NDVI, NDWI, SAVI, TGI, VARI, CI(RedEdge), EVI, GNDVI e NDRE) calculados a partir de imagens multiespectrais adquiridas em seis estádios fenológicos da cultura. As análises revelaram correlações elevadas entre índices associados ao vigor vegetativo e à biomassa (NDVI, SAVI e GNDVI) e entre índices sensíveis ao teor de clorofila (CI(RedEdge) e NDRE). Em contraste, combinações como NDRE–VARI e TGI–CI(RedEdge) apresentaram baixas correlações, indicando independência espectral e potencial complementaridade em modelagens preditivas. Na sequência, foi realizada a caracterização espectral de 251 cultivares de soja, incluindo materiais comerciais e não comerciais, com base nos índices NDRE, TGI e VARI. Os agrupamentos obtidos por meio do algoritmo K-Means permitiram distinguir grupos de cultivares com diferentes padrões fisiológicos e produtivos. O índice NDRE destacou-se pela estabilidade e sensibilidade ao teor de clorofila e ao vigor vegetativo, apresentando relação direta com a produtividade, enquanto os índices TGI e VARI mostraram-se eficientes na identificação das fases iniciais e finais do ciclo da cultura. Por fim, avaliou-se o potencial de integração entre sensoriamento remoto e aprendizado de máquina para a predição da produtividade da soja. O algoritmo Random Forest foi aplicado aos índices NDRE, TGI e VARI, considerando diferentes combinações temporais de coleta. O modelo utilizando todas as datas apresentou $R^2 = 0,629$, enquanto o melhor desempenho foi obtido com apenas duas coletas (62 e 104 dias após o plantio), alcançando $R^2 = 0,7267$, MAE = 96,26 g/parcela e RMSE = 124,09 g/parcela. Esses resultados demonstram que a redução no número de campanhas de voo não compromete a acurácia preditiva quando as coletas são realizadas em estádios fenológicos críticos. De

forma geral, os resultados indicam que os índices NDRE, TGI e VARI capturam dimensões espectrais e fisiológicas complementares da cultura da soja, sendo relevantes para o avanço de metodologias de fenotipagem remota e modelagem espectral. Assim, este trabalho valida o uso de VANTs equipados com sensores multiespectrais como uma ferramenta robusta, reproduzível e cientificamente consistente para o monitoramento e a predição da produtividade da soja, contribuindo para o avanço da agricultura de precisão e dos programas de melhoramento genético baseados em dados.

Palavras-chave: soja; índices espectrais; VANT; Random Forest; fenotipagem remota; produtividade.

ABSTRACT

ALVES, Inglith de Souza Lage, D.Sc., Universidade Federal de Viçosa, November, 2025. **Spectral Analysis and Predictive Modeling for Soybean Crop**. Adviser: Flora Maria de Melo Villar. Co-advisers: Felipe Lopes da Silva and Nerilson Terra Santos.

The use of remote sensing and machine learning techniques has proven fundamental for understanding and predicting the agronomic performance of soybean, particularly in a context of increasing demand for non-destructive and high-precision methods. In this study, the applicability of spectral indices derived from multispectral images acquired by sensors mounted on unmanned aerial vehicles (UAVs) was investigated for remote phenotyping, spectral characterization, and soybean yield prediction. Initially, correlations among nine vegetation indices (NDVI, NDWI, SAVI, TGI, VARI, CI(RedEdge), EVI, GNDVI, and NDRE) calculated from multispectral images acquired at six phenological stages of the crop were analyzed. The analyses revealed strong correlations among indices associated with vegetative vigor and biomass (NDVI, SAVI, and GNDVI), as well as between indices sensitive to chlorophyll content (CI(RedEdge) and NDRE). In contrast, combinations such as NDRE–VARI and TGI–CI(RedEdge) showed low correlations, indicating spectral independence and potential complementarity in predictive modeling. Subsequently, the spectral characterization of 251 soybean cultivars, including both commercial and non-commercial materials, was conducted based on the NDRE, TGI, and VARI indices. Clustering using the K-Means algorithm enabled the identification of groups of cultivars with distinct physiological and productive patterns. The NDRE index stood out due to its stability and sensitivity to chlorophyll content and vegetative vigor, showing a direct relationship with productivity, whereas TGI and VARI proved effective in identifying the early and late stages of the crop cycle. Finally, the potential integration of remote sensing and machine learning for soybean yield prediction was evaluated. The Random Forest algorithm was applied to the NDRE, TGI, and VARI indices using different temporal combinations of data collection. The model using all dates achieved an R^2 of 0.629, whereas the best performance was obtained with only two acquisitions (62 and 104 days after planting), reaching an R^2 of 0.7267, MAE of 96.26 g/plot, and RMSE of 124.09 g/plot. These results demonstrate that reducing the number of flight campaigns does not compromise predictive accuracy when data are collected at critical phenological stages. Overall, the results indicate that the NDRE, TGI, and VARI indices capture complementary spectral and physiological dimensions of soybean crops, contributing to

advances in remote phenotyping and spectral modeling methodologies. Therefore, this study validates the use of UAVs equipped with multispectral sensors as a robust, reproducible, and scientifically consistent tool for soybean monitoring and yield prediction, contributing to the advancement of precision agriculture and data-driven breeding programs.

Keywords: soybean; spectral indices; UAV; Random Forest; remote phenotyping; yield.

LISTA DE FIGURAS

Figura 1.1 – Mapa de localização da área experimental.....	26
Figura 1.2 – Gabarito para plantio.....	27
Figura 1.3 – Exemplo de extração dos dados espectrais a partir da repetição central da cultivar, utilizando média de 10 pixels para cálculo dos índices de vegetação.....	33
Figura 1.4 – Correlação entre pares de índices espectrais ao longo dos dias após o plantio.....	37
Figura 1.5 – Evolução temporal das correlações entre pares de índices espectrais que apresentaram, em pelo menos uma das datas de coleta, valores estatisticamente não significativos	38
Figura 1.6 – Evolução temporal das correlações entre pares de índices espectrais que apresentaram, em pelo menos uma das datas de coleta, valores estatisticamente não significativos	39
Figura 1.7 – Análise de Componentes Principais (PCA) aplicados aos índices espectrais da cultura da soja em seis estádios fenológicos distintos (22, 43, 62, 95, 104 e 123 dias após o plantio – DAP)	41
Figura 2.1 – Variação temporal do Índice VARI para os três clusters de cultivares de soja.....	61
Figura 2.2 – Variação temporal do Índice TGI para os três clusters de cultivares de soja.....	63
Figura 2.3 – Variação temporal do Índice NDRE para os três clusters de cultivares de soja.....	65
Figura 3.1 – Gráfico de dispersão: produtividade real vs. predita para o modelo com 6 datas.....	84
Figura 3.2 – Gráfico de dispersão: modelo com melhor trio de datas. Os valores de RMSE e MAE estão apresentados em gramas por parcela (g/parcela)	86
Figura 3.3 – Gráfico de dispersão: modelo com melhor dupla de datas. Os valores de RMSE e MAE estão apresentados em gramas por parcela (g/parcela)	88

LISTA DE TABELAS

Tabela 1.1. Código de cultivares comerciais utilizadas neste trabalho.....	28
Tabela 1.2. Datas, dias após o plantio e estádios fenológicos utilizados na aquisição das imagens de campo	29
Tabela 1.3. Índices de vegetação selecionados para o estudo.....	32
Tabela 2.1 – Relação entre as datas de coleta de dados, dias de plantio e estágio esperado	53
Tabela 2.2 – Fórmulas e bandas utilizadas para o cálculo dos índices espectrais. NIR (Infravermelho próximo), R (Vermelho), G (Verde), B (Azul) e RE (RedEdge)	56
Tabela 2.3 – Variáveis utilizadas na modelagem espectral da soja	57
Tabela 2.4 – Valores médios e desvios padrão dos índices espectrais por cluster e dias após o plantio	59
Tabela 2.5 – Médias de altura, produtividade e distribuição de ciclos das cultivares por cluster e índice espectral	60
Tabela 3.1. Datas de coleta e estádios fenológicos da soja	80
Tabela 3.2. Relação de IVs calculados. NIR (Infravermelho Próximo), R (Vermelho), G (Verde), B (Azul) e RE (RedEdge)	81
Tabela 3.3. Variáveis utilizadas na modelagem espectral da soja	83
Tabela 3.4. Desempenho do modelo preditivo construído a partir de combinações de três datas de aquisição de imagens, expressos pelos coeficientes de determinação (R^2) e pelos erros médio quadrático (RMSE) e absoluto médio (MAE). Os valores de RMSE e MAE estão apresentados em gramas por parcela (g/parcela)	85
Tabela 3.5. Desempenho do modelo preditivo construído a partir de combinações de duas datas de aquisição de imagens, expressos pelos coeficientes de determinação (R^2) e pelos erros médio quadrático (RMSE) e absoluto médio (MAE). Os valores de RMSE e MAE estão apresentados em gramas por parcela (g/parcela)	87
Tabela 3.6. Comparativo entre os modelos com diferentes números de datas de coleta. Os valores de RMSE e MAE estão apresentados em gramas por parcela (g/parcela).....	89

SUMÁRIO

INTRODUÇÃO GERAL.....	14
DISPOSIÇÃO DO TRABALHO	19
REFERÊNCIAS.....	20
CAPÍTULO 1 – MONITORAMENTO ESPECTRAL DA CULTURA DA SOJA: CORRELAÇÃO ENTRE ÍNDICES DE VEGETAÇÃO PARA CARACTERIZAÇÃO FENOTÍPICA.....	22
RESUMO.....	22
ABSTRACT.....	23
1 INTRODUÇÃO.....	24
2 MATERIAL E MÉTODOS.....	26
3 RESULTADOS.....	35
4 DISCUSSÃO.....	42
5 CONCLUSÃO.....	43
REFERÊNCIAS.....	44
CAPÍTULO 2 – CLASSIFICAÇÃO E ESTUDO ESPECTRAL DE CULTURA DA SOJA.....	49
RESUMO.....	49
ABSTRACT.....	50
1 INTRODUÇÃO.....	51
2 MATERIAL E MÉTODOS.....	52
3 RESULTADOS.....	59
4 DISCUSSÃO.....	66
5 CONCLUSÃO.....	69
REFERÊNCIAS.....	70
CAPÍTULO 3 – APRENDIZADO DE MÁQUINAS APLICADO À PREDIÇÃO DA PRODUTIVIDADE DA CULTURA DA SOJA	74
RESUMO.....	74
ABSTRACT.....	75
1 INTRODUÇÃO.....	76
2 MATERIAL E MÉTODOS.....	78
3 ESULTADOS.....	84
4 DISCUSSÃO.....	89

5 CONCLUSÃO.....	90
REFERÊNCIAS.....	92
CONCLUSÃO GERAL.....	96

INTRODUÇÃO GERAL

O Brasil ocupa posição de destaque no cenário agrícola mundial, respondendo por parcela expressiva das exportações globais e gerando elevado valor econômico e social por meio do cultivo da soja. De acordo com estimativas da Companhia Nacional de Abastecimento (CONAB, 2025), a safra 2024/2025 deve alcançar 169,6 milhões de toneladas, representando um incremento de 14,8% em relação ao ciclo anterior. Esse avanço reflete o contínuo aprimoramento tecnológico e produtivo do setor, impulsionado por práticas de manejo mais eficientes e pelo melhoramento genético das cultivares. No entanto, esse crescimento também evidencia os limites da expansão da fronteira agrícola frente às crescentes exigências ambientais e à necessidade de conservação de ecossistemas nativos. Nesse contexto, biomas como o Cerrado e a Amazônia têm sido fortemente impactados por processos de conversão da vegetação natural associados à expansão da soja e à crescente demanda por novas áreas de cultivo (Mempel et al., 2021).

A produção de soja enfrenta desafios crescentes relacionados à necessidade de aumentar a produtividade sem ampliar proporcionalmente os custos ambientais e socioeconômicos. Diversos estudos destacam que a intensificação agrícola requer estratégias de monitoramento capazes de identificar variações nas lavouras, permitindo detectar precocemente deficiências nutricionais, estresses hídricos, pragas e doenças, fatores que comprometem significativamente o rendimento (Berveglieri et al, 2024).

Além disso, o avanço dos programas de melhoramento genético tem ampliado a demanda por métodos de fenotipagem mais rápidos e precisos. Essas abordagens possibilitam a avaliação em larga escala de características agronômicas e fisiológicas das plantas. Dessa forma, torna-se possível selecionar genótipos mais adaptados às mudanças climáticas, mais eficientes no uso da água e dos nutrientes e com maior estabilidade produtiva em diferentes ambientes (Vymyslicky et al., 2025).

Nesse cenário, o uso de veículos aéreos não tripulados (VANTs) consolidou-se como uma alternativa estratégica para o monitoramento agrícola. Essa tecnologia permite a aquisição de dados com alta resolução espacial e temporal, superando diversas limitações associadas aos sensores orbitais tradicionais. A integração de sensores multiespectrais, hiperespectrais e RGB possibilita a caracterização de atributos fisiológicos do dossel, como teor de clorofila, área foliar e biomassa verde, variáveis essenciais para estimar o desempenho da cultura em diferentes condições de manejo e ambiente (Anzar et al., 2025).

Índices espectrais como o Índice de Diferença Normalizada do Red Edge (NDRE), o Índice Visível Resistente à Atmosfera (VARI) e o Índice de Verde Triangular (TGI) têm se mostrado sensíveis a processos fisiológicos relacionados à fotossíntese, à senescência e a estresses hídricos e nutricionais, constituindo proxies valiosas para análises de vigor e saúde do cultivo. O NDRE, ao incorporar a banda red-edge, apresenta maior capacidade de distinguir variações no teor de nitrogênio e na clorofila foliar, mantendo resposta linear mesmo em estágios de alta biomassa (Rehman et al., 2022). Da mesma forma, índices baseados em bandas RGB, como o VARI e o TGI, demonstram boa correlação com o conteúdo de clorofila e o desenvolvimento do dossel, permitindo estimar o vigor da cultura de forma prática e com baixo custo (Gerardo & Lima, 2023; Panthakkan et al., 2025).

Apesar dos avanços observados no uso de sensores multiespectrais embarcados em veículos aéreos não tripulados (VANTs), persistem desafios substanciais quanto à padronização metodológica, à calibração radiométrica e geométrica dos sensores, e à extrapolação dos resultados para diferentes condições edafoclimáticas e genéticas. Segundo Zhang et al. (2025), a maioria das aplicações ainda depende de procedimentos empíricos e calibrações específicas de campo, o que limita a reprodutibilidade e a transferibilidade dos modelos entre diferentes culturas e ambientes.

A integração de variáveis espectrais a algoritmos de aprendizado de máquina tem se mostrado uma estratégia robusta e promissora para a predição da produtividade da soja, permitindo capturar relações complexas entre os atributos espectrais, estruturais e as respostas agrônômicas da cultura. Conforme demonstrado por Berveglieri et al. (2024), o uso de algoritmos como Random Forest e regressão linear múltipla, alimentados por índices de vegetação e bandas hiperespectrais obtidas por VANTs, alcançou elevadas correlações com os dados observados em campo, com erro médio inferior a 2%. Essa integração amplia o potencial do sensoriamento remoto de alta resolução para modelar a variabilidade espacial da produtividade e oferece subsídios para o aprimoramento de práticas de manejo de precisão e programas de melhoramento genético. Modelos como Random Forest, máquinas de vetores de suporte e redes neurais profundas apresentam elevado potencial para lidar com dados multivariados e colineares (Nill et al., 2022). Apesar desses avanços, a literatura evidencia limitações relevantes na representatividade temporal dos experimentos. A maior parte dos estudos concentra-se em estágios fenológicos tardios, notadamente durante o enchimento de grãos e a maturação fisiológica, que exibem maior correlação direta com a produtividade, mas negligenciam

fases intermediárias do ciclo, como o fechamento do dossel e o florescimento, que poderiam antecipar a estimativa de rendimento. Conforme observado por Shammi et al. (2024), os melhores desempenhos preditivos foram obtidos nos estágios reprodutivos R3 a R5, enquanto as fases vegetativas apresentaram menor estabilidade nas relações espectrais, refletindo a necessidade de ampliar as investigações sobre estágios intermediários para aprimorar modelos de predição de produtividade baseados em sensoriamento remoto. Poucos estudos realizam avaliações comparativas sistemáticas da eficácia de diferentes combinações de índices espectrais, ou quantificam o ganho informativo obtido pela integração de índices derivados de distintas regiões do espectro eletromagnético, como o red edge e o visível. El-Hendawy et al. (2019) demonstraram que não existe um índice espectral universal capaz de representar de forma consistente os parâmetros agronômicos em diferentes genótipos e ambientes, destacando que a acurácia e a estabilidade dos modelos variam amplamente conforme as condições de cultivo. Esses resultados apontam que a transferibilidade e a generalização de modelos espectrais permanecem desafiadoras, já que algoritmos calibrados em uma safra ou localidade frequentemente perdem desempenho quando aplicados a cenários edafoclimáticos ou genéticos contrastantes.

Questões relacionadas à viabilidade operacional do uso de VANTs também permanecem em aberto na literatura. Entre os principais desafios estão a definição do número ótimo de voos ao longo do ciclo da cultura, a frequência ideal de aquisição das imagens e a relação custo-benefício do emprego de sensores com diferentes resoluções espectrais e espaciais. Esses fatores são particularmente relevantes no contexto da agricultura de precisão e dos programas de melhoramento genético, nos quais a eficiência operacional e a escalabilidade das metodologias são aspectos críticos para a adoção dessas tecnologias em larga escala (Rejeb et al., 2022; Ferraz et al., 2024).

Estudos reforçam a relevância da fenotipagem remota, mas também mostram lacunas que ainda limitam sua aplicação em larga escala. Volpato et al. (2021) demonstraram que séries temporais de imagens multiespectrais ou RGB obtidas com apenas três a cinco voos próximos à colheita alcançam acurácia comparável à de séries completas com oito a dez voos (RMSE entre 1,3 e 1,8 dias; R^2 de 0,91–0,94). Esses resultados indicam que a distribuição estratégica das coletas pode otimizar recursos sem perda significativa de informação. No entanto, o estudo concentrou-se na maturação fisiológica (R8), o que mantém em aberto a questão sobre quais estádios fenológicos intermediários são mais informativos para a predição antecipada da produtividade.

De forma complementar, Araus et al. (2018) propuseram um fluxo metodológico baseado em imagens multiespectrais e técnicas de regressão para estimar características morfofisiológicas e a produtividade da soja em diferentes ambientes. Embora os autores tenham relatado desempenho consistente, a abordagem permaneceu restrita a condições controladas, o que limita sua extrapolação para cenários comerciais heterogêneos.

Herrero-Huerta et al. (2020) conduziram uma análise abrangente de índices espectrais derivados de sensores multiespectrais embarcados em veículos aéreos não tripulados, empregando algoritmos de aprendizado de máquina, como Random Forest e XGBoost, para estimar a produtividade da soja. Os autores verificaram que a acurácia dos modelos não está necessariamente relacionada à quantidade de índices empregados, mas à diversidade dos fundamentos espectrais considerados, evidenciando a importância de integrar variáveis provenientes de diferentes faixas do espectro eletromagnético, como o visível, o infravermelho próximo e o red edge. Essa integração favoreceu a representação mais completa das dimensões fisiológicas e biofísicas do dossel e resultou em ganhos expressivos de desempenho preditivo. No entanto, o estudo não contemplou uma análise sistemática da redundância informacional entre índices altamente correlacionados, aspecto que pode comprometer a eficiência estatística e a capacidade de generalização dos modelos de predição de rendimento em condições genéticas e ambientais contrastantes (Herrero-Huerta et al., 2020).

Outros trabalhos também aplicaram índices espectrais combinados a algoritmos de deep learning para classificar acamamento das plantas, estimar a umidade do solo e identificar estádios fenológicos quase em tempo real, como demonstrado por Sarkar et al. (2023). Esses estudos apresentaram elevado potencial de acurácia, mas também evidenciaram limitações críticas relacionadas à generalização dos modelos entre diferentes cultivares, ambientes e safras, o que permanece como um dos principais gargalos da área.

Diante dessas limitações, torna-se necessário investigar quais combinações de datas de aquisição e índices espectrais oferecem maior poder explicativo ao longo do ciclo fenológico da soja. Assim, este trabalho avaliou o potencial de variáveis espectrais obtidas por sensoriamento remoto com VANTs para a predição da produtividade da soja, considerando diferentes combinações de datas de aquisição ao longo do ciclo. O estudo foi estruturado em três etapas: (i) análise das correlações entre nove índices espectrais em distintos estádios de desenvolvimento, com ênfase na identificação de redundâncias e complementaridades; (ii) avaliação da evolução multitemporal dos índices NDRE, TGI e

VARI em cultivares pertencentes a diferentes grupos de maturação e níveis de produtividade; e (iii) modelagem da produtividade da soja a partir de combinações de datas e métricas espectrais, empregando o algoritmo Random Forest para avaliar a assertividade preditiva dos modelos.

DISPOSIÇÃO DO TRABALHO

Esta tese foi estruturada em três capítulos, além da introdução geral e da conclusão final. O Capítulo 1, intitulado “Monitoramento espectral da cultura da soja: correlação entre índices de vegetação para caracterização fenotípica”, teve como objetivo analisar as relações entre nove índices espectrais: NDVI, NDWI, SAVI, TGI, VARI, CI(RedEdge), EVI, GNDVI e NDRE. Os índices foram obtidos por meio de imagens multiespectrais em seis estádios fenológicos da cultura da soja. A análise buscou identificar redundâncias e complementaridades entre os índices, de modo a evidenciar combinações espectralmente independentes com potencial de aplicação em modelagens preditivas e em abordagens de fenotipagem remota.

O Capítulo 2, denominado “Classificação e estudo espectral de cultivares de soja”, abordou a caracterização espectral de 251 cultivares, comerciais e não comerciais, a partir dos índices NDRE, TGI e VARI. Por meio da aplicação do algoritmo de agrupamento K-Means, foram identificados grupos de cultivares com padrões fisiológicos e produtivos distintos, permitindo a correlação entre as respostas espectrais e variáveis agronômicas, como altura de planta e produtividade. Essa análise contribuiu para a compreensão da diversidade espectral entre genótipos.

O Capítulo 3, intitulado “Aprendizado de máquinas aplicado à predição da produtividade da cultura da soja”, avaliou a integração entre dados espectrais multitemporais e algoritmos de aprendizado de máquina, com destaque para o modelo Random Forest. O estudo investigou diferentes combinações temporais de aquisição de imagens, visando determinar os estádios fenológicos mais informativos para a estimativa da produtividade. Foram apresentadas as métricas de desempenho dos modelos, incluindo o coeficiente de determinação (R^2), o erro absoluto médio (MAE) e o erro quadrático médio (RMSE), com o propósito de quantificar a acurácia e a robustez das predições.

REFERÊNCIAS

ANZAR, S. M. et al. Evaluation of uav-based rgb and multispectral vegetation indices for precision agriculture in palm tree cultivation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-G-2025, p.163–170, 2025.

ARAUS, J. et al. Translating high-throughput phenotyping into genetic gain. Trends in Plant Science, v. 23, 03 2018.

BERVEGLIERI, A. et al. Remote prediction of soybean yield using uav-based hyperspectral imaging and machine learning models. AgriEngineering, v. 6, n. 3, p. 3242–3260, 2024.

CONAB – COMPANHIA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira: grãos. v. 12, n. 11, p. 89. Brasília: CONAB, ago. 2025.

EL-HENDAWY, S. et al. Comparative performance of spectral reflectance indices and multivariate modeling for assessing agronomic parameters in advanced spring wheat lines under two contrasting irrigation regimes. Frontiers in Plant Science, v. 10, 11 2019.

FERRAZ, M. et al. Integrating satellite and uav technologies for maize plant height estimation using advanced machine learning. AgriEngineering, v. 6, p. 20–33, 01 2024.

GERARDO, R.; LIMA, I. de. Applying rgb-based vegetation indices obtained from uas imagery for monitoring the rice crop at the field scale: A case study in portugal. Agriculture, v. 13, p. 1916, 09 2023.

Herrero-Huerta, Mónica & Rodríguez-Gonzálvez, Pablo & Rainey, Katy. (2020). Yield prediction by machine learning from UAS-based multi-sensor data fusion in soybean. Plant Methods. 16. 10.1186/s13007-020-00620-6. 2020.

MEMPEL, F.; CORBERA, E. Framing the frontier – tracing issues related to soybean expansion in transnational public spheres. *Global Environmental Change*, v. 69, 06 2021.

NILL, L. et al. Arctic shrub expansion revealed by landsat-derived multitemporal vegetation cover fractions in the western canadian arctic. *Remote Sensing of Environment*, v. 281, p. 113228, 11 2022.

PANTHAKKAN, A. et al. Evaluation of UAV-Based RGB and Multispectral Vegetation Indices for Precision Agriculture in Palm Tree Cultivation. 2025.

REHMAN, T. H.; LUNDY, M. E.; LINQUIST, B. A. Comparative sensitivity of vegetation indices measured via proximal and aerial sensors for assessing n status and predicting grain yield in rice cropping systems. *Remote Sensing*, v. 14, n. 12, 2022.

REJEB, A. et al. Drones in agriculture: A review and bibliometric analysis. *Computers and Electronics in Agriculture*, v. 198, p. 107017, 07 2022.

SARKAR, S. et al. Assessment of soybean lodging using uav imagery and machine learning. *Plants*, v. 12, n. 16, 2023.

SHAMMI, S. A. et al. Application of uav multispectral imaging to monitor soybean growth with yield prediction through machine learning. *Agronomy*, v. 14, p. 672, 03 2024.

VOLPATO, L. et al. Optimization of temporal uas-based imagery analysis to estimate plant maturity date for soybean breeding. *The Plant Phenome Journal*, v. 4, 05 2021.

VYMYSLICK'y, T. et al. Phenotypic characterization of soybean genetic resources at multiple locations: breeding implications for enhancing environmental resilience, yield and protein content. *Frontiers in Plant Science*, Volume 16 - 2025, 2025.

ZHANG, S. et al. A review of the application of uav multispectral remote sensing technology in precision agriculture. *Smart Agricultural Technology*, v. 12, p. 101406, 2025.

CAPÍTULO 1 – MONITORAMENTO ESPECTRAL DA CULTURA DA SOJA: CORRELAÇÃO ENTRE ÍNDICES DE VEGETAÇÃO PARA CARACTERIZAÇÃO FENOTÍPICA

RESUMO

O Brasil consolidou-se como o maior produtor mundial de soja na segunda década do século XXI, impulsionado pelo uso intensivo de tecnologias aplicadas ao sistema produtivo, incluindo avanços em melhoramento genético, mecanização agrícola e manejo de precisão. A estimativa da produtividade, diretamente relacionada ao acúmulo de assimilados fotossintéticos, representa um desafio que exige ferramentas como o sensoriamento remoto. Sensores multiespectrais acoplados a veículos aéreos não tripulados (VANTs) destacam-se por fornecer imagens de alta resolução espacial e temporal, permitindo o cálculo de Índices de Vegetação (IVs), amplamente utilizados para avaliar o desenvolvimento fenológico, a biomassa e a senescência da cultura. Diante disso, objetivou-se com estudo avaliar o comportamento espectral de cultivares comerciais e não comerciais de soja, buscando compreender as relações entre diferentes índices e identificar combinações complementares para aplicações em fenotipagem remota e modelagem preditiva mais precisa. Para tanto, foram utilizadas imagens multiespectrais coletadas em seis períodos distintos ao longo do ciclo da cultura, a partir das quais foram calculados os IVs, NDVI, NDWI, SAVI, TGI, VARI, CI(RedEdge), EVI, GNDVI e NDRE, e analisadas suas correlações entre os diferentes estádios fenológicos. Os resultados evidenciaram uma elevada correlação entre os índices NDVI, SAVI e GNDVI, indicando que esses índices capturam aspectos similares do vigor vegetativo e da biomassa ao longo do ciclo da soja. CI(RedEdge) e NDRE também apresentaram forte associação entre si, reforçando sua sensibilidade conjunta à variação de clorofila. Em contrapartida, NDRE, VARI e TGI exibiram correlações fracas ou negativas, indicando independência espectral e destacando-se como candidatos complementares em modelos multivariados. A identificação de pares com correlações não significativas é essencial para selecionar índices com maior complementaridade fisiológica, favorecendo aplicações em fenotipagem remota e em modelagens preditivas mais robustas e precisas.

Palavras chave: soja, índices espectrais, fenotipagem remota, VANTs.

ABSTRACT

Brazil consolidated itself as the world's largest soybean producer in the second decade of the 21st century, driven by the intensive use of technologies in the production system, including advances in genetic improvement, agricultural mechanization, and precision management. Yield estimation, directly related to the accumulation of photosynthetic assimilates, remains a challenge that requires tools such as remote sensing. Multispectral sensors mounted on unmanned aerial vehicles (UAVs) stand out for providing high spatial and temporal resolution imagery, enabling the calculation of Vegetation Indices (VIs), which are widely used to assess phenological development, biomass, and crop senescence. In this context, the objective of this study was to evaluate the spectral behavior of commercial and non-commercial soybean cultivars, aiming to understand the relationships among different indices and to identify complementary combinations for remote phenotyping applications and more accurate predictive modeling. Multispectral images were collected at six distinct periods throughout the crop cycle, from which the VIs NDVI, NDWI, SAVI, TGI, VARI, CI(RedEdge), EVI, GNDVI, and NDRE were calculated and analyzed across different phenological stages. The results revealed a strong correlation among NDVI, SAVI, and GNDVI, indicating that these indices capture similar aspects of vegetative vigor and biomass throughout the soybean cycle. CI(RedEdge) and NDRE also showed strong association with each other, reinforcing their joint sensitivity to chlorophyll variation. In contrast, NDRE, VARI, and TGI exhibited weak or negative correlations, suggesting spectral independence and standing out as complementary candidates in multivariate models. The identification of index pairs with non-significant correlations is essential to select those with greater physiological complementarity, thus supporting remote phenotyping applications and enabling more robust and accurate predictive modeling.

Keywords: soybean, spectral indices, remote phenotyping, UAVs.

1. INTRODUÇÃO

O Brasil consolidou-se, em 2025, como o maior produtor mundial de soja, com uma safra estimada em 169,6 milhões de toneladas, representando um crescimento de 14,8% em relação ao ciclo anterior. Esse resultado reflete um aumento de 3,2% na área cultivada e de 11,3% na produtividade média (CONAB, 2025). O protagonismo da soja no cenário agrícola mundial reforça a necessidade de estratégias capazes de monitorar a cultura com maior precisão, especialmente diante da crescente demanda por produtividade e sustentabilidade no agronegócio. Nesse contexto, o sensoriamento remoto por veículos aéreos não tripulados (VANTs) tem se destacado como ferramenta fundamental na agricultura de precisão, devido à sua versatilidade técnica e eficiência no monitoramento e diagnóstico de culturas (Guo et al., 2021).

A produtividade da soja está associada ao acúmulo de biomassa e à atividade fotossintética das plantas, processos fortemente influenciados por fatores como manejo, condições climáticas, sanidade e arquitetura do dossel. Estudos indicam que a variabilidade da produção pode ser atribuída em grande parte às condições ambientais, responsáveis por até 60–80% da variação no rendimento, enquanto características genéticas e fisiológicas, como a eficiência fotossintética e a capacidade de fixação biológica de nitrogênio, também desempenham papel central na determinação da produtividade (Kipsshakbayeva et al., 2024).

No entanto, a mensuração direta de variáveis fisiológicas ainda enfrenta limitações práticas, uma vez que os métodos convencionais, geralmente destrutivos e dependentes de análises laboratoriais, são trabalhosos, caros e inviáveis para grandes áreas ou múltiplos pontos amostrais. Revisões recentes destacam que esses métodos fornecem apenas uma visão pontual do campo e comprometem sua aplicação em programas de melhoramento e manejo em larga escala (Islam et al., 2024). Diante dessas limitações, o sensoriamento remoto surge como uma alternativa eficiente e não invasiva, permitindo análises rápidas e de elevada acurácia em diferentes condições de cultivo.

Dentro do espectro de ferramentas de sensoriamento remoto, destacam-se as imagens orbitais e os sistemas embarcados em VANTs. Embora os satélites possuam ampla aplicação no monitoramento agrícola, suas limitações quanto à resolução espacial e à frequência temporal reduzem sua eficácia em áreas menores, como parcelas experimentais. Para superar essas restrições, sensores multiespectrais acoplados a VANTs têm se mostrado altamente eficazes, permitindo a obtenção de imagens em alta

resolução e com flexibilidade temporal, possibilitando o acompanhamento detalhado do desenvolvimento das plantas (Shi et al., 2024). Esses sensores captam dados nas faixas do visível e do infravermelho próximo (NIR), que podem ser convertidos em índices de vegetação (IVs), traduzindo a resposta espectral da vegetação em métricas quantitativas e informativas (Miller et al., 2019).

A diversidade de índices de vegetação (IVs) disponíveis permite sua classificação conforme a função fisiológica que cada um melhor representa. O Índice de Vegetação da Diferença Normalizada (NDVI), o Índice de Vegetação Ajustado ao Solo (SAVI) e o Índice de Vegetação por Diferença Normalizada do Verde (GNDVI) estão associados ao vigor vegetativo e à biomassa, apresentando alta correlação com a estrutura e o acúmulo de fitomassa, como evidenciado em estudos recentes (Israel et al., 2025). O Índice de Diferença Normalizada da Água (NDWI) está diretamente relacionado ao conteúdo hídrico e tem sido amplamente utilizado na identificação de estresse hídrico em culturas agrícolas (Caturegli et al., 2020). Já o Índice de Vegetação Melhorado (EVI) oferece vantagens em ambientes de alta biomassa, pois reduz interferências atmosféricas e do solo, mantendo sensibilidade estrutural mesmo em cobertura densa (Montero et al., 2023).

Para diagnóstico nutricional, o Índice de Diferença Normalizada do Red-Edge (NDRE) e o Índice de Clorofila – Red-Edge (CI-RedEdge) destacam-se por sua sensibilidade à clorofila e à detecção precoce de estresse fisiológico, sendo apontados como eficazes na avaliação de deficiência nutricional (Zhang et al., 2024). O Índice de Verde Triangular (TGI) tem sido utilizado para estimar senescência e maturação foliar, apresentando resultados consistentes em estudos com imagens aéreas de alta resolução (Polat et al., 2024). Já o Índice Visível Resistente à Atmosfera (VARI), que depende apenas das bandas visíveis, representa uma alternativa prática para monitoramento em diferentes condições atmosféricas, apresentando boa resposta em ambientes urbanos e agrícolas (Chahbi et al., 2025).

Apesar da ampla disponibilidade desses índices, muitos estudos ainda os aplicam simultaneamente sem considerar as correlações existentes entre eles, o que pode gerar redundância informacional e reduzir a eficiência de modelos preditivos (SILVA JUNIOR et al., 2018). Pesquisas recentes reforçam que análises mais criteriosas, considerando o comportamento espectral dos índices ao longo dos diferentes estádios fenológicos, são fundamentais para aprimorar a eficácia de modelos de monitoramento agrícola. Além disso, avanços metodológicos demonstram que índices derivados de sensores multiespectrais e de imagens RGB obtidas por VANTs podem ser utilizados como

alternativa viável para acompanhar o desenvolvimento das culturas, embora com menor precisão em relação aos índices tradicionais (Da Silva et al., 2024).

Diante desse cenário, o presente estudo teve como objetivo analisar a correlação entre nove índices de vegetação obtidos por imagens multiespectrais captadas por VANTs ao longo de seis estádios fenológicos da soja, identificando aqueles que fornecem informações complementares e não redundantes, contribuindo para a fenotipagem remota e para a construção de modelos preditivos mais assertivos.

2. MATERIAL E MÉTODOS

2.1. Área Experimental

O experimento foi conduzido no campo experimental Prof. Diogo Alves de Melo ($20^{\circ}46'10.9''$ S, $42^{\circ}52'13.8''$ W), localizado na Universidade Federal de Viçosa (UFV), em Viçosa, Minas Gerais (Figura 1.1). Nesta área, foram plantadas cinco cultivares comerciais de soja e 251 cultivares não comerciais no dia 28 de outubro de 2022. Para garantir a correta identificação e manejo das diferentes cultivares durante o experimento, foi projetado um gabarito específico para o plantio (Figura 1.1), assegurando a disposição organizada das cultivares dentro do campo experimental.

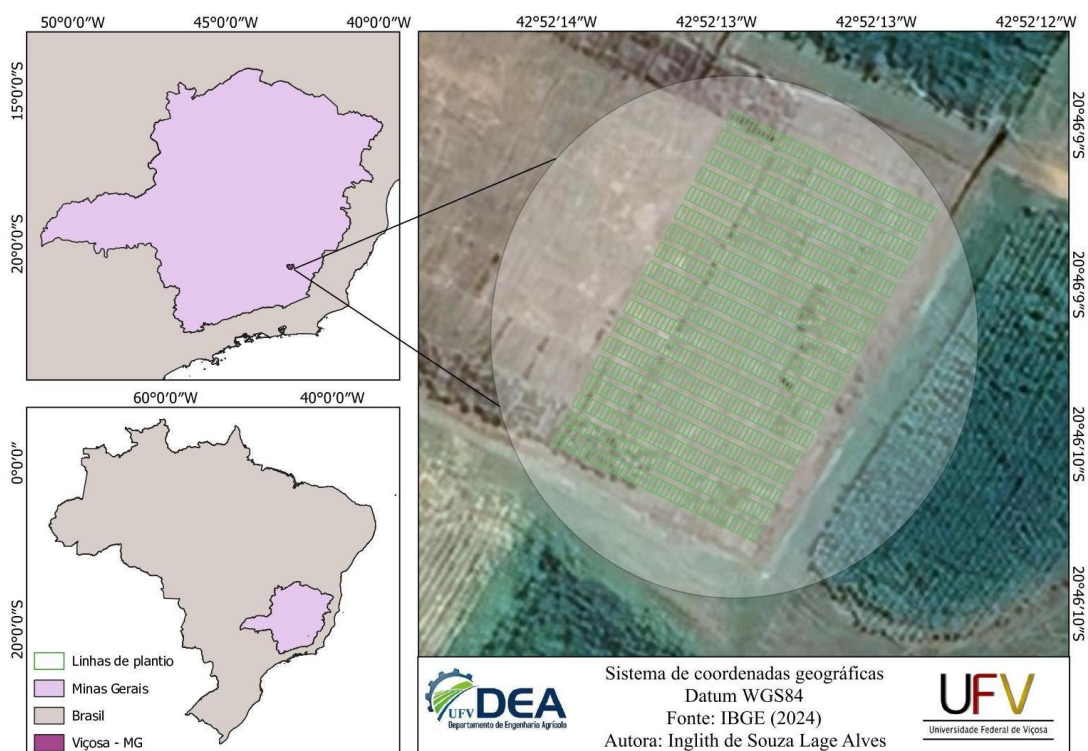


Figura 1.1. Mapa de localização da área experimental.

A área destinada ao experimento foi dividida em dois blocos, sendo o bloco 1 o principal foco deste estudo. Nesse bloco, foram estabelecidas 21 linhas de plantio, com espaçamento de 0,7 m entre linhas e largura de parcela de 1,5 m. A primeira e a última linhas foram ocupadas por cultivares de bordadura, utilizadas para reduzir os efeitos de borda sobre as demais cultivares avaliadas. As cinco cultivares comerciais, ilustradas na Figura 1.2, foram representadas pelas cores azul, amarelo, vermelho, laranja e verde. As demais cultivares, que estão em fase de análise e ainda não são comerciais, foram identificadas por códigos. Cada cultivar, tanto comercial quanto não comercial, foi plantada três vezes, em sequência, ao longo da linha, para minimizar a influência das plantas vizinhas nas informações coletadas em cada cultivar.

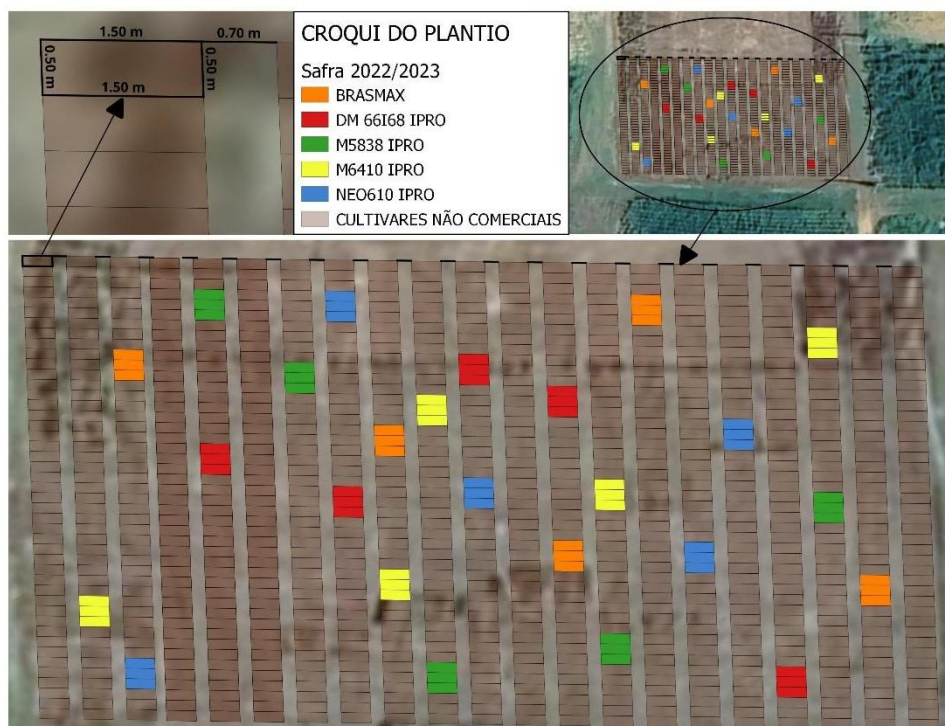


Figura 1.2. Gabarito para plantio.

As cinco cultivares comerciais foram: BRASMAX Cromo TF IPRO, M5838 IPRO, NEO610 IPRO, M6410 IPRO e DM 66I68 IPRO. As cultivares eram pertencentes ao Banco de Germoplasma do Programa de Melhoramento de Soja do Departamento de Agronomia da Universidade Federal de Viçosa, que foram utilizadas como genitores. O acompanhamento dos ensaios foi realizado semanalmente, com foco na identificação e no controle de pragas e doenças. As práticas de manejo adotadas seguiram as orientações técnicas específicas para a cultura da soja, conforme as necessidades observadas em cada

área experimental (SILVA et al., 2022). As cultivares comerciais (Tabela 1.1) foram codificadas dentro do experimento com os respectivos identificadores.

Tabela 1.1. Código de cultivares comerciais utilizadas neste trabalho

	CÓDIGO 1	CÓDIGO 2	CÓDIGO 3	CÓDIGO 4	CÓDIGO 5
BRASMAX	5017	5095	5164	5204	5265
M5838 IPRO	5004	5073	5112	5183	5249
NEO610 IPRO	5024	5088	5146	5214	5246
M6410 IPRO	5030	5105	5134	5196	5239
DM 66I68 IPRO	5048	5064	5138	5189	5260

2.2. Coleta de dados fotogramétricos e pontos de controle

A coleta de dados fotogramétricos foi realizada utilizando um veículo aéreo não tripulado (VANT) do modelo Matrice 100 (Fabricado por SZ DJI Technology Co., Ltd., Shenzhen, Guangdong, China). O modelo do VANT facilita a integração de diferentes sensores, tornando-se ideal para aplicações em sensoriamento remoto. O sensor multiespectral integrado foi o RedEdge-MX (fabricado pela MicaSense, Inc., Seattle, Washington, EUA). Esse equipamento é composto por cinco sensores CMOS independentes, que capturam imagens em distintos comprimentos de onda: azul (475 nm, largura de banda de 20 nm), verde (560 nm, 20 nm), vermelho (668 nm, 10 nm), RedEdge (717 nm, 10 nm) e infravermelho próximo (840 nm, 40 nm).

A aquisição das imagens foi realizada em seis datas distintas ao longo do ciclo da cultura (Tabela 1.2). A escolha desses períodos teve como objetivo capturar a dinâmica do crescimento da soja, abrangendo diferentes estádios fenológicos da cultura. Os voos ocorreram, em cada data de coleta, entre 11:30 e 12:30 horas, em condições de céu limpo. A escolha do horário de voo está relacionada à condição de maior proporção de radiação direta sobre a superfície da Terra e menores erros de reflectância (HONKAVAARA et al., 2009). Os planos de voos foram programados com altura de 20 metros, 75% de sobreposição lateral e longitudinal. Além das imagens capturadas no momento do voo, adicionalmente foram coletadas imagens do painel de calibração radiométrica do sensor RedEdge-MX.

Em ambos os experimentos foram distribuídos, nas extremidades do talhão, 8 Pontos de Controle (GCPs). Estes são componentes fundamentais para o mapeamento com veículos aéreos não tripulados (VANTs), pois proporcionam maior precisão espacial,

correção de distorções e integração eficiente dos dados obtidos com outras fontes geoespaciais (CHO et al., 2023).

Os GCPs foram distribuídos estrategicamente na área experimental e marcados com alvos de padrão xadrez nas cores preta e branca, medindo 0,30 x 0,30 metros, confeccionados em lona do tipo front light.

As coordenadas geográficas dos pontos de controle foram obtidas utilizando receptores GNSS RTK modelo Topomap TPS T10 (fabricado por Topomap Nova Lima, Brasil). Os dois receptores GNSS foram empregados com funções distintas: um atuando como estação base fixa e o outro como receptor móvel (rover). A estação base foi posicionada em um local estável dentro da área de estudo, com o registro contínuo de suas coordenadas durante a coleta, permitindo a aplicação de correções posteriores. O receptor móvel, por sua vez, foi utilizado para coletar as coordenadas dos 8 GCPs. A solução RTK foi incluída no levantamento pois pode atingir o nível centimétrico ou mesmo milimétrico quando as ambiguidades de fase da portadora são corretamente corrigidas (Wang et al., 2021).

Tabela 1.2. Datas, dias após o plantio e estádios fenológicos utilizados na aquisição das imagens de campo

Data	Dias após o plantio	Ilustração da cultura	Estádio fenológico
19/11/2022	22		V3/V4
11/12/2022	43		R1/R2
29/12/2022	62		R3/R4
31/01/2023	95		R5/R6

09/02/2023	104		R6/R7
28/02/2023	123		R8

2.3. Processamento de dados

Após a coleta das imagens, o processamento dos dados multiespectrais foi realizado no software Agisoft Metashape Professional (Versão 1.6.3, Agisoft LLC, St. Petersburg, Rússia), conforme o fluxo de trabalho recomendado para sensores MicaSense, como o RedEdge-MX. O procedimento seguiu as etapas descritas segundo o tutorial oficial da Agisoft (2024), que detalha o processamento desde a importação das imagens até a exportação dos produtos finais, incluindo a calibração radiométrica por painel de refletância.

Inicialmente, todas as imagens capturadas, incluindo as do painel de calibração, foram adicionadas ao projeto como sistema multicâmera. O software foi configurado para detectar automaticamente as imagens do painel, utilizando metadados para organizá-las em pastas específicas e aplicar máscaras que isolam apenas o painel de calibração. Em casos nos quais essa detecção automática não ocorreu, a segmentação das imagens foi feita manualmente. A calibração da refletância foi realizada com base nos valores fornecidos pelo certificado do painel. As opções “Use reflectance panels” e “Use sun sensor” foram ativadas para garantir a correção radiométrica adequada, de acordo com as especificações do sensor MicaSense RedEdge-MX (AGISOFT, 2024).

Posteriormente, as imagens foram alinhadas para geração da nuvem de pontos e cálculo das posições relativas das câmeras. Com essa estrutura preliminar, procedeu-se à inserção dos Pontos de Controle Terrestre (GCPs), obtidos por levantamento com receptores GNSS, que antes de iniciar a geração dos mosaicos, foram processados por meio do serviço de Processamento de Ponto Preciso (PPP), disponibilizado pelo IBGE, adotando o sistema de referência geodésico SIRGAS 2000 (IBGE, 2023). Os GCPs foram importados no formato CSV e marcados manualmente nas imagens correspondentes. Essa etapa possibilitou o georreferenciamento do modelo com base em coordenadas mais precisas do terreno. O modelo foi otimizado utilizando os parâmetros de calibração da

câmera, seguido pela construção da nuvem de pontos densa, modelo digital de elevação (DEM) e ortomosaico. A geração do ortomosaico utilizou o DEM como superfície de referência e o modo de mistura de imagens foi desabilitado para evitar interpolações indesejadas (AGISOFT, 2024). Os ortomosaicos foram gerados e salvos na extensão .tiff, apresentando resolução espacial média de 1,24 cm/pixel.

2.4. Geração de Índices de Vegetação

Para o cálculo dos índices de vegetação, foi desenvolvido um script em Python executado diretamente no console Python do QGIS (versão 3.28.5, QGIS Development Team, 2023), aproveitando as funcionalidades da API PyQGIS. O código foi projetado para automatizar o processamento e minimizar a necessidade de intervenção manual, otimizando o fluxo de trabalho e a reprodutibilidade dos resultados.

As imagens multiespectrais, exportadas como camadas raster individuais, por banda, estavam nomeadas com prefixos padronizados, (b1, b2, b3, b4 e b5), representando, respectivamente, as bandas azul, verde, vermelha, infravermelho próximo (NIR) e RedEdge. O script foi programado para percorrer o diretório onde as imagens estavam armazenadas, identificar os arquivos de acordo com esses prefixos e os organizar em listas correspondentes.

Para cada fórmula dos índices de vegetação, Tabela 1.3, foi realizado o cálculo pixel a pixel com base nas bandas de entrada. Os resultados foram exportados em arquivos GeoTIFF georreferenciados e os nomes dos arquivos gerados incluíam o nome do índice e um sufixo identificador do dia da coleta de campo, facilitando a organização. Esse processo foi repetido para cada voo, permitindo a geração automática dos resultados e eliminando a necessidade de processamento manual repetitivo. Os índices escolhidos para a execução da metodologia estão apresentados na Tabela 1.3.

Tabela 1.3. Índices de vegetação selecionados para o estudo

Índice de vegetação	Equação	Referência
Índice de Vegetação da Diferença Normalizada	$NDVI = \frac{NIR - R}{NIR + R}$	(Rouse et al., 1973)
Índice de Diferença Normalizada da Água	$NDWI = \frac{G - NIR}{G + NIR}$	(Mc Ferters, 1996)
Índice de Vegetação Melhorado	$EVI = 2,5 \left(\frac{NIR - R}{NIR + 6 * R - 7,5 * B + 1} \right)$	(Huete et al., 1997)
Índice de Vegetação Ajustado ao Solo	$SAVI = \frac{NIR - R}{(NIR + R + 0,5) * (1,0 + 0,5)}$	(Huete, 1988)
Índice Visível Resistente à Atmosfera	$VARI = \frac{G - R}{G + R - B}$	(Gitelson et al., 2002)
Índice de Diferença Normalizada do Red Edge	$NDRE = \frac{NIR - RE}{NIR + RE}$	(Barnes et al., 2000)
Índice de Clorofila (RedEdge)	$CI_{RedEdge} = \frac{NIR}{RE} - 1$	(Pan et al., 2023)(Gitelson et al., 2003)
Índice de Vegetação por Diferença Normalizada do Verde	$GNDVI = \frac{NIR - G}{NIR + G}$	Gitelson et al., 1996
Índice de Verde Triangular	$TGI = G - 0,39 * R - 0,61 * B$	(Lemes et al., 2022)(Hunt et al., 2011)
NIR (Infravermelho próximo), R (Vermelho), G (Verde), B (Azul) e RE (RedEdge) ¹		

Os nove índices de vegetação utilizados neste estudo foram calculados a partir dos ortomosaicos multiespectrais processados. Para cada cultivar, foi extraído o valor médio com base em 10 amostragens pontuais de pixel, representando a reflectância da vegetação.

Cada cultivar foi plantada em três repetições no campo experimental (Figura 1.3), organizadas de forma a minimizar a interferência entre genótipos adjacentes. Para garantir a independência espectral das medições, os valores foram extraídos exclusivamente da repetição central de cada cultivar, evitando bordas laterais e possíveis efeitos de contaminação entre parcelas.

A amostragem foi realizada manualmente no ambiente QGIS, por meio da ferramenta “*Point Sampling Tool*”, selecionando 10 pontos dentro da área ocupada pela vegetação da cultivar, com atenção para evitar pixels mistos com solo exposto ou

1

bordaduras. O valor médio desses 10 pixels foi considerado representativo da resposta espectral da cultivar para cada índice em cada data de voo.

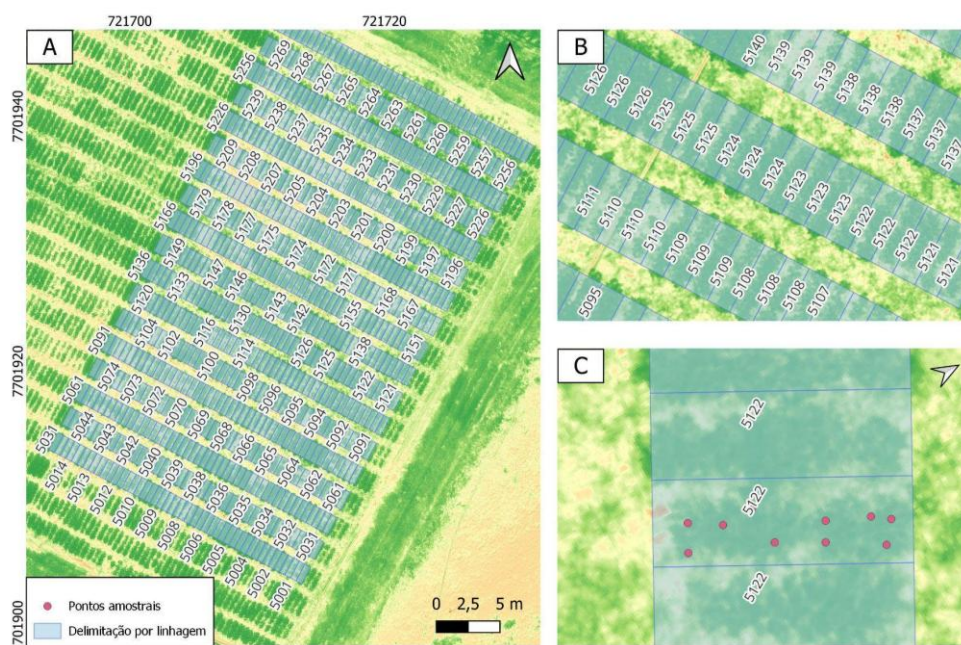


Figura 1.3. Exemplo de extração dos dados espectrais a partir da repetição central da cultivar, utilizando média de 10 pixels para cálculo dos índices de vegetação.

2.5. Análise de Correlação dos Índices Espectrais

A relação entre os índices espectrais ao longo do ciclo da soja foi investigada por meio de uma análise de correlação. Para cada dia avaliado, realizou-se uma análise de correlação para cada par de variáveis. Além disso, a plotagem da análise estatística em um gráfico de uma análise de componentes principais (PCA) também foi realizada.

A Análise de Componentes Principais (Principal Component Analysis – PCA) é uma técnica estatística multivariada amplamente utilizada para explorar a estrutura de covariância entre variáveis. Seu principal objetivo é reduzir a dimensionalidade de conjuntos de dados complexos, projetando-os em um espaço de menor dimensão sem perda significativa da informação essencial (Hongyu et al., 2016). O procedimento consiste em transformar linearmente variáveis originalmente correlacionadas em um novo conjunto de eixos ortogonais, denominados componentes principais (PCs), cuja

quantidade depende do conjunto analisado. O primeiro componente principal explica a maior parte da variabilidade das variáveis, enquanto o segundo, independente do primeiro, explica a segunda maior variação, e assim sucessivamente. Em geral, espera-se que os dois primeiros componentes expliquem pelo menos 80% da variância acumulada (Jolliffe; Cadima, 2016).

Cada componente é associado a autovetores, que refletem a contribuição relativa das variáveis originais. Valores próximos de +1,0 ou -1,0 indicam alta relevância da variável na composição do componente, sendo essas relações normalmente expressas por coeficientes de correlação de Pearson. O coeficiente de correlação de Pearson (r) foi empregado para quantificar a intensidade e a direção da relação linear entre variáveis contínuas. Consequentemente, uma matriz de correlação foi gerada para cada uma das seis datas avaliadas, englobando todas as combinações possíveis entre os índices.

2.5.1 Teste de Significância e Normalização

A significância estatística das correlações foi avaliada por meio de um teste de hipótese aplicado ao coeficiente de correlação de Pearson (r). A hipótese nula ($H_0: \rho = 0$) assumiu a inexistência de relação linear entre as variáveis analisadas, enquanto a hipótese alternativa ($H_1: \rho \neq 0$) postulou a existência de uma correlação linear estatisticamente significativa. (Obilor & Amadi, 2018).

Considerando o tamanho amostral do estudo ($n = 251$), o teste foi conduzido com 249 graus de liberdade ($gl = n - 2$), utilizando um nível de significância de 5% ($\alpha = 0,05$) em um teste bicaudal. Com base nas tabelas da distribuição t de Student, o valor crítico de t para 249 graus de liberdade foi de aproximadamente $\pm 1,970$. O valor crítico correspondente ao coeficiente de correlação de Pearson foi obtido a partir da relação entre t e r , conforme as equações (1) e (2).

$$t = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} \quad (1)$$

$$r_{crítico} = \frac{t}{\sqrt{t^2 + (n-2)}} = \frac{1,970}{\sqrt{1,970^2 + (251 - 2)}} = 1,238 \quad (2)$$

Em que:

t = valor calculado da estatística t ;

n = tamanho amostral ($n = 251$);

$r_{\text{crítico}}$ = valor crítico mínimo do coeficiente de correlação.

Dessa forma, o intervalo de não rejeição da hipótese nula foi definido como $[-0,1238, +0,1238]$. Coeficientes de correlação (r) situados dentro desse intervalo foram considerados estatisticamente não significativos, enquanto valores fora dele indicaram correlação significativa ao nível de 5% ($p < 0,05$).

Além disso, para evitar redundância nas análises, foi adotado um processo de normalização das combinações entre pares de índices espectrais. Devido à simetria do coeficiente de correlação ($r_{xy} = r_{yx}$), pares como "CI-TGI" e "TGI-CI" foram tratados como equivalentes e organizados em ordem alfabética, garantindo unicidade e consistência nos resultados.

2.5.2 Representação Gráfica e Análise de Frequência

A interpretação dos resultados foi complementada por representações gráficas, elaboradas para facilitar a visualização. Foram gerados quatro tipos de gráficos:

- **Todas as correlações par a par:** ilustrando por meio de matrizes todas as correlações;
- **Pares mais estáveis:** Focando nas combinações cujas correlações se mantiveram consistentemente dentro da região de aceitação ao longo de todos os dias analisados.

3. RESULTADOS

3.1 Correlações Médias entre Índices Espectrais

A análise das correlações entre os índices espectrais ao longo do ciclo fenológico da soja apresentou variações na intensidade das associações lineares entre os pares avaliados. O coeficiente de correlação de Pearson (r) foi utilizado como métrica de dependência, considerando-se as médias calculadas em seis épocas distintas de coleta: 22, 43, 62, 95, 104 e 123 dias após a semeadura. A Figura 1.4 apresenta a matriz de correlação obtida para todas as combinações de índices.

Entre os pares avaliados, a combinação entre o índice de clorofila (CI) e o índice de diferença normalizada do Red Edge (NDRE) apresentou a correlação média mais elevada observada no estudo ($r = 0,952$). Esse resultado evidencia uma associação positiva e consistente ao longo do ciclo da cultura, além da sua semelhança matemática da equação. Na literatura, estudos com sensoriamento remoto mostram que índices como CI e outros derivados da região do red edge são sensíveis ao teor de clorofila e à estrutura do dossel (Shi et al., 2023). Esses atributos estão diretamente relacionados ao vigor fisiológico das plantas e sustentam a escolha de um deste para modelagem preditiva.

Na sequência, observou-se correlações elevadas entre o par NDVI–SAVI, o que reflete a semelhança espectral entre ambos, uma vez que esses índices derivam das mesmas bandas do vermelho e do infravermelho próximo e respondem de forma semelhante às variações de biomassa e cobertura do dossel. Em experimentos com soja, Da Silva et al. (2024) também observaram que NDVI e SAVI apresentaram desempenho superior na predição da produtividade, evidenciando alta similaridade entre esses índices e sua relação direta com o rendimento da cultura.

A relação entre NDWI e TGI destaca a complementaridade fisiológica entre índices associados ao conteúdo de umidade da vegetação e à intensidade da pigmentação verde foliar. Na cultura da soja, variações na disponibilidade hídrica afetam simultaneamente o teor de água nas folhas e a concentração de clorofila, o que altera a reflectância nas regiões do infravermelho próximo e do visível. Estudos demonstram que o aumento do estresse hídrico, seja por déficit ou excesso de água, reduz o conteúdo de clorofila, a eficiência fotossintética e modifica a resposta espectral da planta (Shi et al., 2024; Adegoye et al., 2023). Essa interação explica a alta correlação entre NDWI e TGI, uma vez que ambos os índices respondem a processos fisiológicos interligados, em que o status hídrico influencia diretamente a manutenção dos pigmentos e, consequentemente, o vigor do dossel.

Em contraste, alguns pares apresentaram correlações médias muito baixas ou até negativas, a exemplo de NDWI–CI ($r = -0,096$) e NDWI–VARI ($r = -0,075$). Esses resultados apontam que índices relacionados à umidade do dossel, como o NDWI, não apresentam comportamento linear consistente com índices voltados à detecção de clorofila (CI) ou à pigmentação foliar em bandas do visível (VARI), sobretudo ao longo das diferentes fases fenológicas. A ausência de correlação significativa sugere que tais índices capturam atributos fisiológicos distintos e devem ser empregados de forma complementar, e não redundante, em abordagens de monitoramento espectral. Essa

interpretação é apresentada em estudos que demonstram que índices baseados no visível, como VARI e TGI, são particularmente sensíveis à pigmentação verde, enquanto índices hídricos, como NDWI, refletem propriedades relacionadas ao teor de água do dossel, justificando a baixa associação linear entre eles (Chahbi et al., 2025).

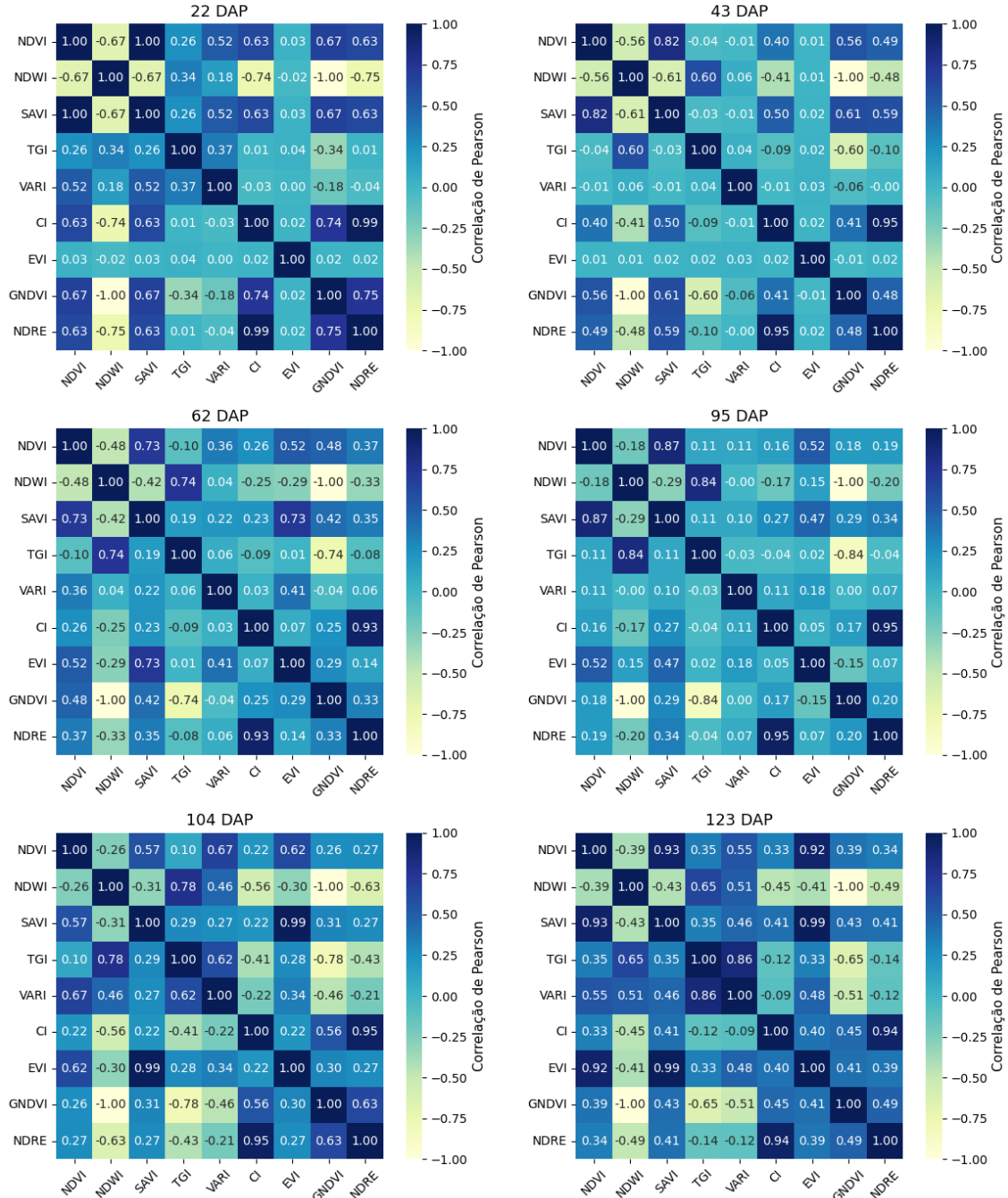


Figura 1.4. Correlação entre pares de índices espectrais ao longo dos dias após o plantio.

A aplicação do teste de hipóteses possibilitou classificar os pares em dois grupos: aqueles com correlações estatisticamente significativas ao longo de todas as datas de

coleta; e aqueles cuja associação não se mostrou significativa em pelo menos um dos períodos avaliados. A Figura 1.5 ilustra o resultado das combinações que apresentaram baixas correlações em determinadas épocas, enquanto a Figura 1.6 evidencia os pares que permaneceram, por mais dias, dentro da região de não significância estatística.

Entre os pares recorrentes em valores não significativos destacaram-se CI-VARI, NDRE-VARI e TGI-CI, que não apresentaram associações estatisticamente relevantes em cinco dos seis períodos analisados. Essas combinações podem indicar independência espectral relevante para modelagens complementares, reduzindo redundâncias. Entretanto, análises adicionais revelaram que NDRE-VARI e CI-VARI apresentaram padrões de correlação bastante semelhantes ao longo de todas as datas, sugerindo comportamento espectral próximo entre essas combinações. Tal achado motivou a avaliação direta entre NDRE e CI, cuja correlação mostrou-se elevada, confirmando que, apesar das interações fracas com o VARI, esses dois índices compartilham informações espectrais relevantes.

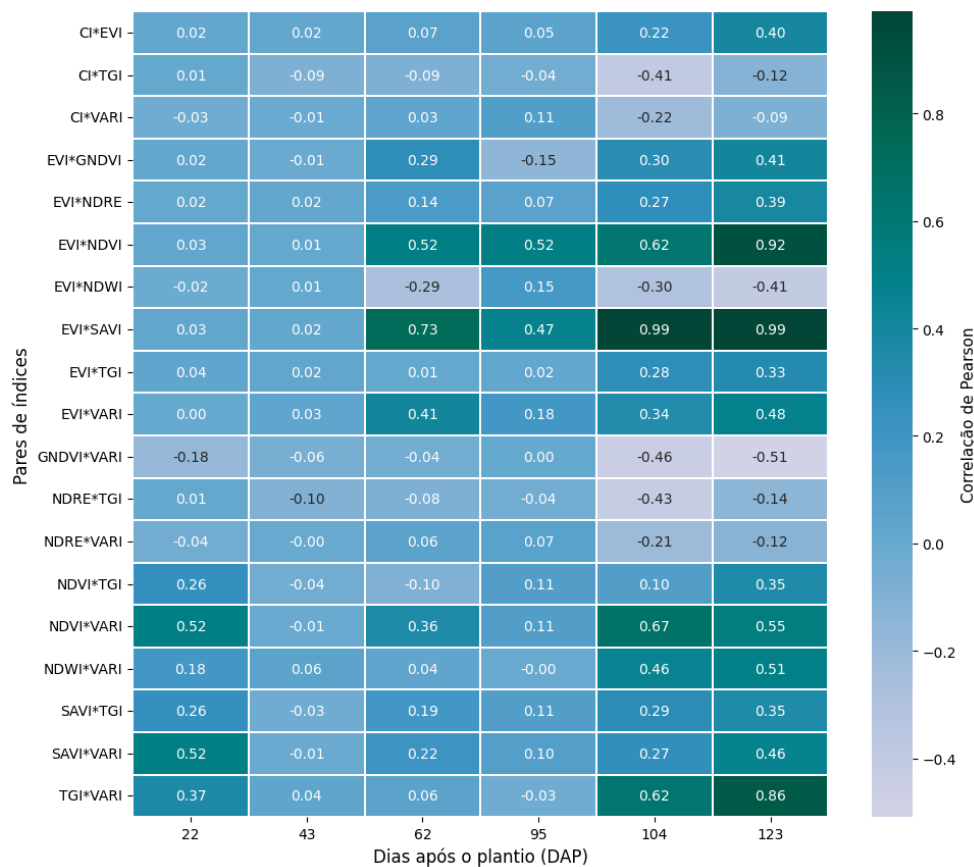


Figura 1.5. Evolução temporal das correlações entre pares de índices espectrais que

apresentaram, em pelo menos uma das datas de coleta, valores estatisticamente não significativos.

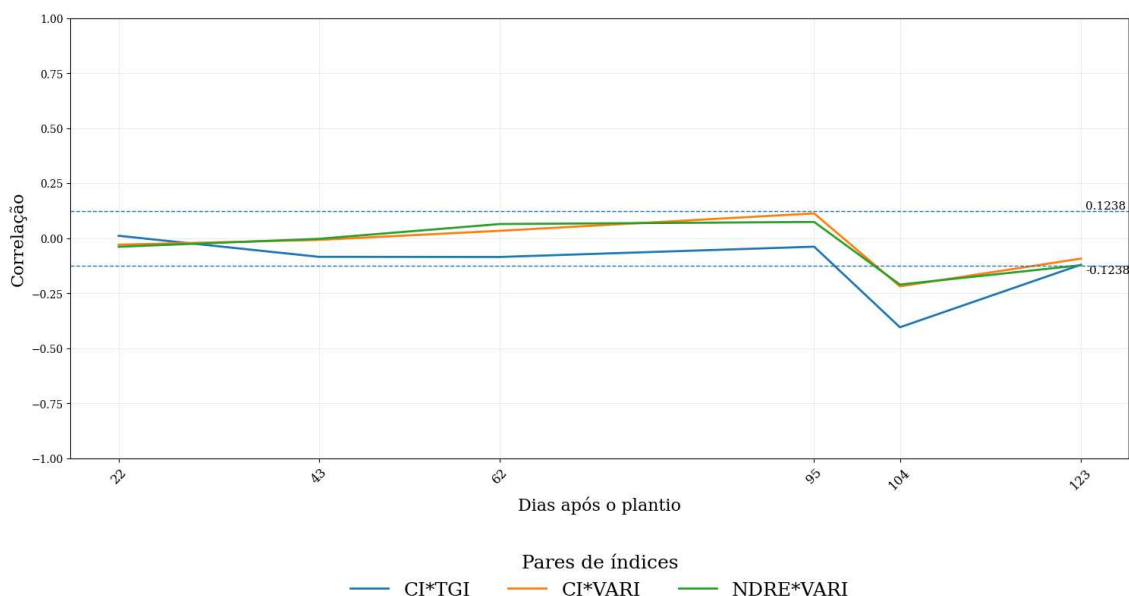


Figura 1.6. combinações cujas correlações se mantiveram consistentemente dentro da região de aceitação ao longo de todos os dias analisados.

3.2 Análise de Componentes Principais dos Índices Espectrais ao Longo do Ciclo da Soja

A análise de Componentes Principais (PCA), Figura 1.7, possibilitou a identificação de padrões entre os índices espectrais ao longo das diferentes datas avaliadas. Em todas as datas avaliadas, os dois primeiros componentes principais explicaram entre 60,6% e 75,0% da variância total, sendo o PC1 responsável pela maior proporção, variando de 34,7% a 52,6%, e o PC2 entre 16,7% e 29,5%. De modo geral, o PC1 representou o gradiente de vigor vegetativo e acúmulo de biomassa, com predominância dos índices NDVI, NDRE e GNDVI, enquanto o PC2 expressou a variação associada também a clorofila e ao teor hídrico, com maior contribuição de TGI, VARI e NDWI. Nos estádios iniciais (22 e 43 DAP), observou-se maior dispersão das cultivares e vetores mais longos, indicando elevada variabilidade espectral entre genótipos em fase de estabelecimento. Nesses estádios, o EVI apresentou vetores curtos e baixa contribuição para a variância total, refletindo sua menor sensibilidade nas fases iniciais. Em 62 DAP, houve redução acentuada da dispersão dos pontos e encurtamento

dos vetores, demonstrando alta colinearidade entre os índices. A partir de 95 DAP, observou-se aumento expressivo na dispersão das cultivares e alongamento dos vetores, indicando maior variabilidade espectral entre os genótipos e intensificação das diferenças fisiológicas associadas ao avanço para os estádios reprodutivos. Na fase final (123 DAP), a soma da variância explicada pelos dois primeiros componentes principais foi de 71,5%, indicando alta capacidade da PCA em sintetizar a variabilidade espectral. Observa-se redução na dispersão das cultivares e encurtamento dos vetores, evidenciando menor heterogeneidade espectral entre os genótipos em decorrência da senescência foliar e do declínio do teor de clorofila. Nesse estágio, o VARI e o TGI apresentaram vetores convergentes e curtos, apontando o nivelamento das diferenças de pigmentação e vigor do dossel.

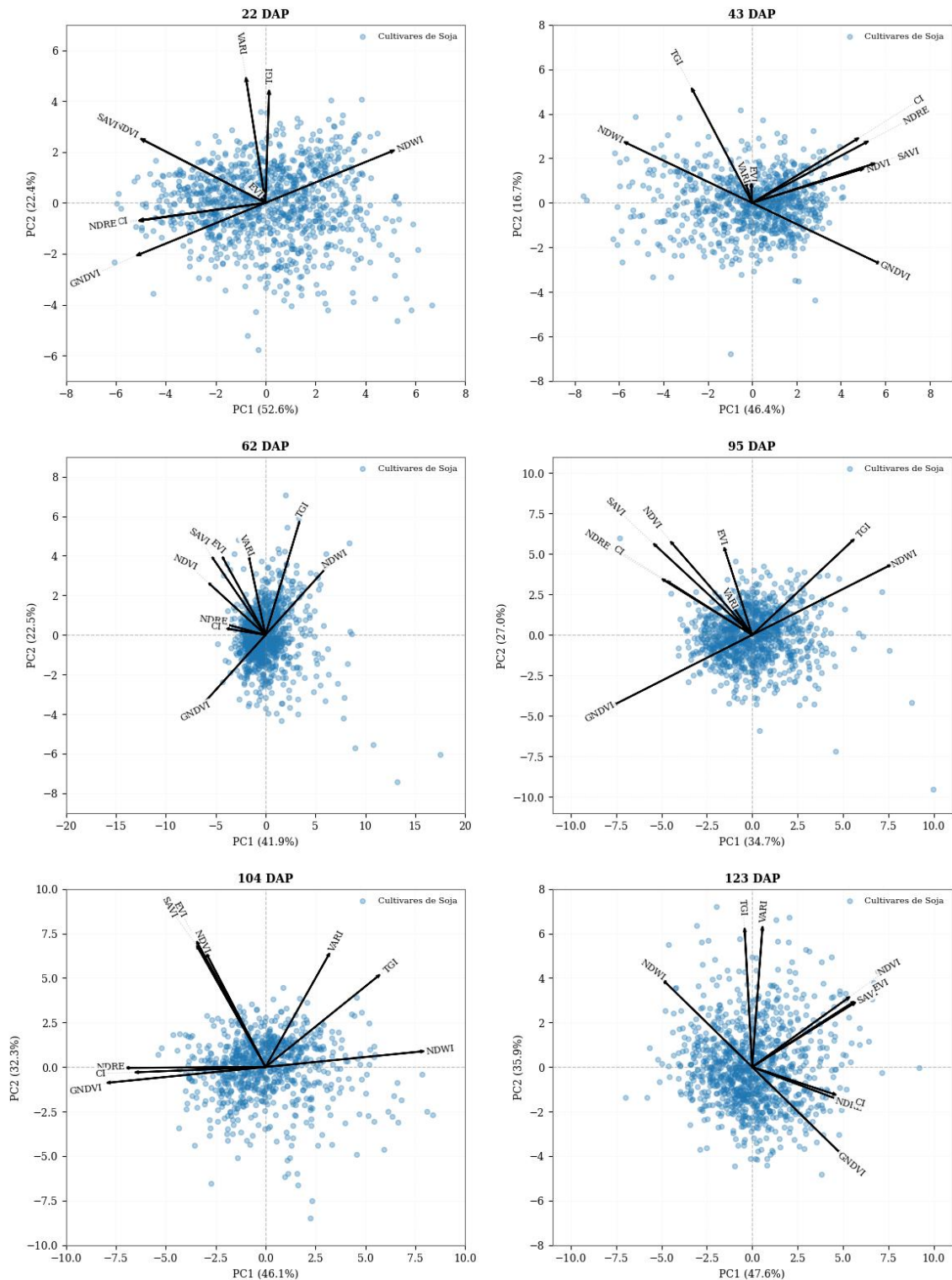


Figura 1.7. Análise de Componentes Principais (PCA) aplicados aos índices espectrais da cultura da soja em seis estádios fenológicos distintos (22, 43, 62, 95, 104 e 123 dias após o plantio – DAP). Os pontos representam as observações correspondentes às diferentes cultivares de soja, enquanto os vetores indicam a magnitude e a direção da contribuição de cada índice para a variância total explicada pelos dois primeiros componentes principais (PC1 e PC2).

4. DISCUSSÃO

Na análise de correlações entre índices espectrais se destacaram dois grandes grupos de comportamento: pares altamente associados, cuja resposta espectral tende a ser redundante, e pares com correlação fraca, que capturam dimensões fisiológicas distintas da cultura da soja.

4.1. Índices com Maior Correlação

Tanto o CI quanto o NDRE apresentaram comportamento espectral semelhante ao longo do ciclo, refletindo respostas paralelas às variações no conteúdo de clorofila, na estrutura foliar e semelhança equacional. Estudos anteriores confirmam que índices baseados na banda red-edge são altamente sensíveis à atividade fotossintética e à concentração de pigmentos foliares (Pan et al., 2023).

Em seguida, NDVI-SAVI apresentou alta correlação, em conformidade com estudos que demonstram o acoplamento entre índices de vigor e biomassa e a mitigação do efeito do solo pelo SAVI (Amankulova et al., 2023; Shammi et al., 2024). De forma complementar, NDWI-TGI mostrou correlação positiva moderada ao longo do ciclo, coerente com a sensibilidade do NDWI ao status hídrico e do TGI às variações de pigmentação verde e conteúdo de clorofila (Yang et al., 2024; Pan et al., 2023).

4.2 Índices com Menor Correlação

Pares que envolvem o NDWI e índices associados à clorofila ou à pigmentação na faixa do visível apresentaram correlações fracas ou negativas, como NDWI-CI ($r = -0,096$) e NDWI-VARI ($r = -0,075$). Essa independência estatística está de acordo com o comportamento espectral descrito por Huang et al. (2022), que demonstraram que os índices relacionados ao conteúdo hídrico do dossel e aqueles vinculados à clorofila representam propriedades biofísicas distintas no mapeamento de áreas de soja. De forma semelhante, Shi et al. (2023) confirmaram que os índices espectrais sensíveis às regiões red-edge e visível estão fortemente correlacionados ao teor de clorofila foliar ao longo dos estádios de crescimento, sendo minimamente influenciados pelas variações de umidade do dossel. Esses resultados reforçam a independência espectral entre o NDWI

(sensível à umidade) e os índices relacionados aos pigmentos (atividade fotossintética e intensidade de coloração verde).

Pares como CI–VARI, NDRE–VARI e TGI–CI apresentaram correlações fracas na maior parte das datas, indicando que índices sensíveis à clorofila e índices RGB de pigmentação, intensidade do verde, capturam dimensões fisiológicas distintas e complementares. Resultados em soja obtidos por Randelović et al. (2023) demonstram correlações fracas entre o TGI e índices de clorofila, evidenciando que esses grupos espectrais descrevem diferentes aspectos do dossel. De forma consistente, Lemes et al. (2022) caracterizaram o TGI como um indicador de pigmentação e tonalidade verde, reforçando que índices RGB e red-edge representam propriedades espectralmente independentes associadas, respectivamente, à aparência foliar e ao teor de clorofila.

4. CONCLUSÃO

Os resultados evidenciam que os índices NDRE, TGI e VARI apresentam comportamentos espectrais distintos ao longo do ciclo da cultura da soja, configurando-se como variáveis complementares na caracterização do dossel. As análises de correlação indicaram baixa associação entre alguns desses índices, sugerindo que cada um deles captura diferentes padrões de resposta espectral da cultura. A integração desses índices em abordagens multiespectrais amplia a capacidade de representação da variabilidade espectral observada ao longo do desenvolvimento da soja, fornecendo subsídios consistentes para a construção de modelos preditivos mais robustos.

REFERÊNCIAS

ADEGOYE, G. A.; OLORUNWA, O. J.; ALSAJRI, F. A.; WALNE, C. H.; WIJEWANDANA, C.; KETHIREDDY, S. R.; REDDY, K. N.; REDDY, K. R. **Waterlogging Effects on Soybean Physiology and Hyperspectral Reflectance during the Reproductive Stage.** Agriculture, v.13, p.844, 2023.

AGISOFT. **MicaSense RedEdge MX processing workflow (including Reflectance Calibration) in Agisoft Metashape Professional.** Modificado em: 12 nov. 2024. Disponível em: <https://agisoft.freshdesk.com/support/solutions/articles/31000148780-micasense-rededge-mx-processing-workflow-including-reflectance-calibration-in-agisoft-metashape-pro>. Acesso em: 17 set. 2025.

AMANKULOVA, K.; FARMONOV, N.; AKRAMOVA, P.; TURSUNOV, I.; MUCSI, L. **Comparison of PlanetScope, Sentinel-2, and Landsat 8 data in soybean yield estimation within-field variability with random forest regression.** Heliyon, v.9, n.6, e17432, 2023.

BARNES, E. M.; CLARKE, T. R.; RICHARDS, S. E.; COLAIZZI, P. D.; HABERLAND, J.; KOSTRZEWSKI, M.; WALLER, P.; CHOI, C.; RILEY, E.; THOMPSON, T.; LASCANO, R. J.; LI, H.; MORAN, M. S. **Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data.** Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA, p.16-19, 2000.

CATUREGLI, L.; MATTEOLI, S.; GAETANI, M. et al. **Effects of water stress on spectral reflectance of bermudagrass.** Sci Rep, v.10, p.15055, 2020.

CHAHBI, M.; EL GANADI, Y.; IDRISSE GARTOUMI, K. **Monitoring urban vegetation by GeoAI driven multi-scale indices: a case study of Rabat, Morocco.** J. Umm Al-Qura Univ. Eng. Archit., v.16, p.649–663, 2025.

CHO, J. M.; LEE, B. K. **Plano de Utilização de GCP e PPK para Lidar com a Interrupção do Sinal RTK em Fotogrametria RTK-UAV.** Drones, v.7, p.265, 2023.

CONAB – COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira: grãos.** v. 12, n. 11, p. 89. Brasília: CONAB, ago. 2025.

DA SILVA, A. A.; SILVA, F. C. S.; GUIMARÃES, C. M.; SALEH, I. A.; CRUS NETO, J. F.; EL-TAYEB, M. A.; ABDEL-MAKSoud, M. A.; AGUILERA, J. G.; ABDELGAWAD, H.; ZUFFO, A. M. **Spectral indices with different spatial resolutions in recognizing soybean phenology.** PLoS ONE, v.19, n.9, e0305610, 2024.

GITELSON, A. A.; GRITZ, Y.; MERZLYAK, M. N. **Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves.** J. Plant Physiol., v. 160, p. 271–282, 2003.

GITELSON, A. A.; KAUFMAN, Y. J.; MERZLYAK, M. N. **Use of a green channel in remote sensing of global vegetation from EOS-MODIS.** Remote Sensing of Environment, v. 58, n. 3, p. 289–298, 1996.

GITELSON, A.; KAUFMAN, Y.; STARK, R.; RUNDQUIST, D. **Novel Algorithms for Remote Estimation of Vegetation Fraction.** Remote Sensing of Environment, v. 80, p. 76-87, 2002.

GUO, W.; CARROLL, M. E.; SINGH, A.; SWETNAM, T. L.; MERCHANT, N.; SARKAR, S.; SINGH, A. K.; GANAPATHYSUBRAMANIAN, B. **UAS-based plant phenotyping for research and breeding applications.** Plant Phenomics, 2021.

HONGYU, K.; SANDANIELO, V. L. M.; OLIVEIRA JUNIOR, G. J. **Análise de Componentes Principais: Resumo Teórico, Aplicação e Interpretação.** E&S Engineering and Science, v.5, n.1, p.83–90, 2016.

HONKAVAARA, E.; ARBIOL, R.; MARKELIN, L.; MARTINEZ, L.; CRAMER, M.; BOVET, S.; CHANDELIER, L.; ILVES, R.; KLONUS, S.; MARSHAL, P.; SCHLÄPFER, D.; TABOR, M.; THOM, C.; VEJE, N. **Digital airborne photogrammetry - a new tool for quantitative remote sensing? A state-of-the-art review on radiometric aspects of digital photogrammetric images.** Remote Sensing, v.1, n.3, p.577–605, 2009.

HUANG, Y.; QIU, B.; CHEN, C.; ZHU, X.; WU, W.; JIANG, F.; LIN, D.; PENG, Y. **Automated soybean mapping based on canopy water content and chlorophyll content using Sentinel-2 images.** International Journal of Applied Earth Observation and Geoinformation, v.109, 2022.

HUETE, A. R. **A soil-adjusted vegetation index. Remote Sensing of Environment**, Elsevier Science Publishing Co., New York, USA. v. 25, p. 295-309, 1988.

HUETE, A. R. et al. **A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS. Remote Sens. Environ**, n.59, p.440-451, 1997.

HUNT, E.; DAUGHTRY, C.; EITEL, J.; LONG, D. **Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index. Agronomy Journal**, v.103, p.1090, 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **IBGE-PPP: serviço on-line para pós-processamento de dados GNSS – manual do usuário**. Rio de Janeiro: IBGE, Diretoria de Geociências, Coordenação de Geodésia, jun. 2023

ISLAM, M.; BIJJAHALLI, S.; FAHEY, T. et al. **Destructive and non-destructive measurement approaches and the application of AI models in precision agriculture: a review. Precision Agric**, v.25, p.1127–1180, 2024.

ISRAEL, R.; AKINTUNDE-ALO, D. A.; OLUWAJUWON, T. V. **Comparative evaluation of Landsat 9 vegetation indices for estimating urban forest biomass in the botanical garden of University of Ibadan, Nigeria. Discov. For.**, v.1, p.32, 2025.

JOLLIFFE, I. T.; CADIMA, J. **Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci**, v.374, n.2065, 2016.

KIPSHAKBAYEVA, G. et al. **Assessment of the genetic parameters of soybean genotypes for precocity and productivity in the various cultivation conditions. Heliyon**, v. 10, e36135, 2024.

LEMES, E. M.; COELHO, L.; ANDRADE, S. L. D.; OLIVEIRA, A. D. S.; MARQUES, M. G.; NASCIMENTO, F. M. A. D.; CUNHA, J. P. A. R. D. **Triangular Greenness Index to Evaluate the Effects of Dicamba in Soybean. AgriEngineering**, v.4, p.758-769, 2022.

MCFEETERS, S. K. **The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing**, v.17, n.7, p.1425-1432, 1996.

MILLER, J. J.; SCHEPERS, J. S.; SHAPIRO, C. A.; ARNESON, N. J.; ESKRIDGE, K. M.; OLIVEIRA, M. C.; GIESLER, L. J. **Characterizing soybean vigor and productivity using multiple crop canopy sensor readings.** Field Crops Research, v.216, p.22-31, 2018.

MONTERO, D.; AYBAR, C.; MAHECHA, M. D. et al. **A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research.** Sci Data, v.10, p.197, 2023.

OBILO, E. I.; AMADI, E. **Test for Significance of Pearson's Correlation Coefficient (r).** International Journal of Innovative Mathematics, Statistics & Energy Policies, v.6, n.1, p.11-23, Jan-Mar, 2018.

PAN, D.; LI, C.; YANG, G.; REN, P.; MA, Y.; CHEN, W.; FENG, H.; CHEN, R.; CHEN, X.; LI, H. **Identification of the Initial Anthesis of Soybean Varieties Based on UAV Multispectral Time-Series Images.** Remote Sens., v.15, p.5413, 2023.

POLAT, N.; MEMDUHOĞLU, A.; KAYA, Y. **Triangular Greenness Index Analysis for Monitoring Fungal Disease in Pine Trees: A UAV-based Approach.** Bartın Orman Fakültesi Dergisi, v.26, n.2, p.1-15, 2024.

RANDELOVIĆ, P.; ĐORĐEVIĆ, V.; MILADINOVIĆ, J. et al. **High-throughput phenotyping for non-destructive estimation of soybean fresh biomass using a machine learning model and temporal UAV data.** Plant Methods, v.19, p.89, 2023.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W.; FREDEN, S. C. **Monitoring vegetation systems in the Great Plains with ERTS.** Proceedings of 3rd Earth Resources Technology Satellite-1 Symposium. 1973.

SHAMMI, S. A.; HUANG, Y.; FENG, G.; TEWOLDE, H.; ZHANG, X.; JENKINS, J.; SHANKLE, M. **Application of UAV Multispectral Imaging to Monitor Soybean Growth with Yield Prediction through Machine Learning.** Agronomy, v.14, p.672, 2024.

SHI, H.; GUO, J.; AN, J.; TANG, Z.; WANG, X.; LI, W.; ZHAO, X.; JIN, L.; XIANG, Y.; LI, Z. et al. **Estimation of Chlorophyll Content in Soybean Crop at Different Growth Stages Based on Optimal Spectral Index.** Agronomy, v.13, p.663, 2023.

SHI, H.; LIU, Z.; LI, S.; JIN, M.; TANG, Z.; SUN, T.; LIU, X.; LI, Z.; ZHANG, F.; XIANG, Y. **Monitoring Soybean Soil Moisture Content Based on UAV Multispectral and Thermal-Infrared Remote-Sensing Information Fusion.** *Plants (Basel)*, v.13, n.17, 2024.

SHI, W.; LI, Y.; ZHANG, W.; YU, C.; ZHAO, C.; QIU, J. **Monitoring and zoning soybean maturity using UAV remote sensing.** *Industrial Crops and Products*, v.222, 2024.

SILVA JUNIOR, C. A.; NANNI, M. R.; SHAKIR, M.; TEODORO, P. E.; OLIVEIRA-JUNIOR, J. F.; CEZAR, E.; GOIS, G.; LIMA, M.; WOJCIECHOWSKI, J. C.; SHIRATSUCHI, L. S. **Soybean varieties discrimination using non-imaging hyperspectral sensor.** *Infrared Physics & Technology*, v.89, p.338–350, 2018.

SILVA, E. E.; BAIO, F. H. R.; RIBEIRO TEODORO, L. P.; SILVA JUNIOR, C. A.; BORGES, R. S.; TEODORO, P. E. **UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation.** *Remote Sensing Applications: Society and Environment*, v.18, 2020.

WANG, J.; XU, T.; NIE, W.; XU, G. **A simplified processing algorithm for multi-baseline RTK positioning in urban environments.** *Measurement*, v.179, p.109446, 2021.

YANG, W.; LI, Z.; CHEN, G.; CUI, S.; WU, Y.; LIU, X.; MENG, W.; LIU, Y.; HE, J.; LIU, D. et al. **Soybean (Glycine max L.) Leaf Moisture Estimation Based on Multisource Unmanned Aerial Vehicle Image Feature Fusion.** *Plants*, v.13, p.1498, 2024.

ZHANG, J.; HU, Y.; LI, F.; FUE, K. G.; YU, K. **Meta-Analysis Assessing Potential of Drone Remote Sensing in Estimating Plant Traits Related to Nitrogen Use Efficiency.** *Remote Sens.*, v.16, p.838, 2024.

CAPÍTULO 2 – CLASSIFICAÇÃO E ESTUDO ESPECTRAL DA CULTURA DA SOJA

RESUMO

O monitoramento espectral de cultivares de soja constitui uma abordagem robusta e precisa para a fenotipagem remota e a identificação de genótipos com desempenho agrônomo superior. Este estudo teve como objetivo caracterizar o comportamento espectral de 251 cultivares de soja, comerciais e não comerciais, e avaliar a capacidade discriminante dos índices NDRE (Normalized Difference Red Edge Index), TGI (Triangular Greenness Index) e VARI (Visible Atmospherically Resistant Index) na diferenciação de grupos com distintas características fenológicas e produtivas. As imagens multiespectrais foram adquiridas por veículo aéreo não tripulado (VANT) em seis estádios fenológicos representativos do ciclo da cultura, e os agrupamentos foram determinados pelo algoritmo K-Means, considerando variáveis espectrais, morfológicas e agrônomicas. Os resultados indicaram que o NDRE foi o índice mais sensível e fisiologicamente consistente para representar o acúmulo de clorofila e o vigor vegetativo, apresentando correlação direta com a produtividade. Os grupos de cultivares com valores elevados de NDRE entre 64 e 104 dias após o plantio refletiram maior eficiência fotossintética e acúmulo de biomassa, resultando nas maiores produtividades médias (1157,7 g/parcela), especialmente em cultivares de ciclo médio. O TGI demonstrou elevada capacidade de detecção da senescência foliar, enquanto o VARI apresentou maior sensibilidade ao vigor vegetativo inicial, revelando-se complementares na caracterização das diferentes fases do desenvolvimento da cultura. De modo abrangente, a análise espectral multitemporal permitiu distinguir padrões fisiológicos e fenológicos entre os grupos de cultivares, consolidando o sensoriamento remoto com câmeras multiespectrais embarcadas em VANT como uma ferramenta robusta, não destrutiva e de alta reprodutibilidade para o monitoramento do desenvolvimento da soja, a seleção genotípica e o manejo de precisão.

Palavras-chave: cultivares de soja; imagens multiespectrais; índices de vegetação; fenotipagem remota.

ABSTRACT

Spectral monitoring of soybean cultivars represents a robust and precise approach for remote phenotyping and for the identification of genotypes with superior agronomic performance. This study aimed to characterize the spectral behavior of 251 soybean cultivars, both commercial and non-commercial, and to evaluate the discriminant capacity of the NDRE (Normalized Difference Red Edge Index), TGI (Triangular Greenness Index), and VARI (Visible Atmospherically Resistant Index) in differentiating groups with distinct phenological and productive characteristics. Multispectral images were acquired using an unmanned aerial vehicle (UAV) at six representative phenological stages of the crop cycle, and clustering was performed using the K-Means algorithm, considering spectral, morphological, and agronomic variables. The results indicated that NDRE was the most sensitive and physiologically consistent index to represent chlorophyll accumulation and vegetative vigor, showing a direct correlation with yield. Cultivar groups exhibiting high NDRE values between 64 and 104 days after planting reflected greater photosynthetic efficiency and biomass accumulation, resulting in the highest mean yields (1157.7 g per plot), particularly in medium-cycle cultivars. The TGI demonstrated strong capability for detecting leaf senescence, while VARI showed higher sensitivity to early vegetative vigor, proving complementary in characterizing different phases of crop development. Overall, the multitemporal spectral analysis enabled the distinction of physiological and phenological patterns among cultivar groups, confirming that remote sensing with UAV-mounted multispectral cameras is a robust, non-destructive, and highly reproducible tool for monitoring soybean development, supporting genotype selection, and advancing precision management strategies.

Keywords: soybean cultivars; multispectral imaging; vegetation indices; remote phenotyping.

1. INTRODUÇÃO

A soja (*Glycine max* L.) é uma das culturas agrícolas mais importantes do mundo, exercendo papel central na economia global e na segurança alimentar. Sua relevância ultrapassa a produção de grãos, abrangendo subprodutos destinados à alimentação humana, animal e à indústria, o que contribui para a geração de empregos e o desenvolvimento regional, especialmente em países produtores como o Brasil. Segundo a Companhia Nacional de Abastecimento (CONAB, 2025), o Brasil tornou-se o maior produtor mundial da leguminosa, com uma produção estimada em 169,6 milhões de toneladas, representando cerca de 40% da produção global. Esse desempenho é impulsionado, em grande parte, pelos avanços no melhoramento genético, que ampliaram a adaptação de cultivares a diferentes condições ambientais e sistemas de manejo (Silva et al., 2024).

Nos últimos anos, o avanço das tecnologias digitais na agricultura consolidou o uso de ferramentas de monitoramento baseadas em sensoriamento remoto, especialmente os veículos aéreos não tripulados (VANTs) equipados com câmeras multiespectrais. Essas tecnologias permitem a obtenção de dados em alta resolução espacial e temporal, viabilizando o acompanhamento detalhado do desenvolvimento das lavouras, a identificação precoce de estresses e a tomada de decisões mais assertivas no manejo das culturas (Aldrighettoni; D'Urso, 2025). Nesse contexto, os índices de vegetação (IVs) surgem como métricas fundamentais para a análise do estado fisiológico das plantas de forma não destrutiva, contínua e escalável.

Entre os índices espectrais aplicados ao monitoramento da soja, destacam-se o Índice de Diferença Normalizada do Red Edge (NDRE), o Índice de Verde Triangular (TGI) e o Índice Visível Resistente à Atmosfera (VARI). O NDRE utiliza a banda red edge, sensível à concentração de clorofila, sendo eficaz em diagnósticos nutricionais e na detecção precoce de estresses fisiológicos, como deficiência hídrica e desequilíbrios nutricionais (Li et al., 2022). De forma complementar, o Índice de Clorofila (CI-RedEdge) também explora essa banda, aprofundando a análise do teor de clorofila. O TGI, por sua vez, é amplamente utilizado na estimativa da concentração de clorofila e no acompanhamento da senescência foliar, sendo particularmente útil nas fases finais do ciclo da cultura (Da Silva et al., 2022). Já o VARI, que depende apenas das bandas visíveis, fornece estimativas de cobertura vegetal menos sensíveis a variações

atmosféricas e de iluminação, o que o torna adequado para ambientes heterogêneos e em condições de campo variáveis (Anzar et al., 2025).

Para interpretar adequadamente os dados obtidos por meio desses índices, é essencial compreender o ciclo fenológico da soja, que influencia diretamente o comportamento espectral da cultura. O ciclo é dividido em estádios vegetativos e reprodutivos, ambos cruciais para o desempenho agrônômico da planta (Fehr et al., 1977). Nos estádios vegetativos (V1, V2, V3), ocorre o desenvolvimento da estrutura foliar e radicular, determinante para a interceptação de luz e a absorção de nutrientes e água. Já os estádios reprodutivos (R1 a R8) abrangem desde o início da floração até a maturação fisiológica, sendo altamente sensíveis a estresses ambientais. Durante essa fase, índices como o TGI permitem monitorar a senescência e, indiretamente, o status hídrico da planta, auxiliando na definição de momentos críticos para intervenções no manejo (Ferraz et al., 2024).

Adicionalmente, a resposta espectral da cultura pode variar entre diferentes grupos genéticos. A distinção entre cultivares comerciais e não comerciais é estratégica para o avanço do cultivo da soja. Enquanto as cultivares comerciais se destacam pela estabilidade e alta produtividade, as não comerciais, frequentemente experimentais, representam uma fonte valiosa de variabilidade genética, podendo expressar características desejáveis, como resistência a estresses bióticos e abióticos e maior adaptação a ambientes específicos (Mendonça et al., 2022). Nesse sentido, a aplicação do sensoriamento remoto, combinada a algoritmos de aprendizado de máquina, tem se mostrado eficaz para diferenciar cultivares em campo, com base em assinaturas espectrais e características morfofisiológicas (Guo et al., 2021).

Diante desse cenário, este capítulo tem como objetivo explorar o potencial dos índices espectrais NDRE, TGI e VARI no monitoramento multitemporal da soja, avaliando sua capacidade de diferenciar grupos de cultivares a partir de características fenológicas e produtivas. A proposta busca contribuir com ferramentas de apoio à seleção genotípica e ao manejo de precisão, favorecendo maior eficiência e sustentabilidade na produção da cultura.

2. MATERIAL E MÉTODOS

2.1. Área Experimental e Delineamento

O experimento foi conduzido no Campo Experimental Prof. Diogo Alves de Melo (20°46'10.9" S, 42°52'13.8" W), pertencente à Universidade Federal de Viçosa (UFV), em Viçosa, Minas Gerais. Nessa área foram cultivadas cinco cultivares comerciais de soja e 251 cultivares não comerciais, com plantio realizado em 28 de outubro de 2022. Para assegurar a identificação e o manejo adequados das cultivares ao longo do experimento, foi elaborado um gabarito específico de plantio. A área de plantio foi dividida em dois blocos, sendo o Bloco 1 o principal foco deste estudo. Nesse bloco, foram estabelecidas 21 linhas de plantio, cada uma com 1,5 metro de largura e 0,7 metro de espaçamento entre linhas. As linhas externas foram ocupadas por cultivares de bordadura, utilizadas exclusivamente para minimizar os efeitos de borda sobre as parcelas de interesse. Cada parcela experimental foi composta por três linhas de plantio com 1,5 m de comprimento, espaçadas em 0,5 m entre si, totalizando área de 2,25 m² por parcela.

As cinco cultivares comerciais foram BRASMAX Cromo TF IPRO, M5838 IPRO, NEO610 IPRO, M6410 IPRO e DM 66I68 IPRO. As demais cultivares, em fase experimental e ainda não comercializadas, foram identificadas por códigos alfanuméricos. Cada cultivar, comercial ou não comercial, foi semeada em triplicata ao longo das linhas de plantio, com o objetivo de reduzir a interferência entre plantas vizinhas e aumentar a precisão das observações espectrais e agronômicas.

As cultivares utilizadas pertencem ao Banco de Germoplasma do Programa de Melhoramento de Soja do Departamento de Agronomia da UFV, sendo empregadas como genitores em cruzamentos experimentais. O acompanhamento da lavoura foi realizado semanalmente, com foco no monitoramento fitossanitário e no controle de pragas e doenças. As práticas de manejo seguiram as recomendações técnicas específicas para a cultura da soja, sendo ajustadas conforme as condições observadas durante o ciclo (SILVA et al., 2022).

Ao longo do ciclo da cultura, foram coletadas informações sobre produtividade, altura na floração, data de floração e classificação do ciclo fenológico (precoce, médio e tardio). As coletas foram realizadas de forma sistemática, conforme apresentado na Tabela 2.1, abrangendo diferentes estádios de desenvolvimento da cultura, o que possibilitou a análise multitemporal do comportamento espectral e agrônomo das cultivares.

Tabela 2.1. Relação entre as datas de coleta de dados, dias de plantio e estágio esperado

Evento	Data	Dias desde o plantio	Estádio esperado
Plantio	28/10/2022	0	V1

Coleta 1	19/11/2022	22	V3/V4
Coleta 2	10/12/2022	43	R1/R2
Coleta 3	29/12/2022	62	R3/R4
Coleta 4	31/01/2023	95	R5/R6
Coleta 5	09/02/2023	104	R6/R7
Coleta 6	28/02/2023	123	R8

2.2. Aquisição e processamento de dados multiespectrais

A aquisição dos dados fotogramétricos foi conduzida com o uso de um veículo aéreo não tripulado (VANT) Matrice 100 (SZ DJI Technology Co., Ltd., Shenzhen, China), configurado para integrar diferentes sensores. O sensor multiespectral embarcado foi o RedEdge-MX (MicaSense Inc., Seattle, EUA), composto por cinco sensores CMOS independentes que capturam as faixas espectrais azul (475, 10 nm), verde (560, 10 nm), vermelho (668, 5 nm), RedEdge (717, 5 nm) e infravermelho próximo (840, 20 nm). As imagens foram coletadas em seis datas distintas ao longo do ciclo fenológico da soja, entre 11:30 h e 12:30 h, sob condições de céu claro, visando minimizar variações na reflectância (HONKAVAARA et al., 2009). Os voos foram planejados a 20 m de altitude, com 75% de sobreposição lateral e longitudinal, acompanhados do registro de imagens do painel de calibração radiométrica para posterior correção das reflectâncias.

Foram instalados oito Pontos de Controle em Terra (GCPs) nas extremidades do talhão, dispostos estrategicamente para garantir precisão geométrica e integração espacial dos produtos gerados (CHO et al., 2023). Os GCPs, confeccionados em lona front light no padrão xadrez preto e branco (0,30 × 0,30 m), tiveram suas coordenadas obtidas por meio de receptor GNSS RTK Topomap TPS T10 (Topomap, Brasil), operando em modo base e rover. A utilização da correção RTK assegurou acurácia centimétrica a milimétrica, conforme descrito por Wang et al. (2021).

Após a etapa de aquisição das imagens, o processamento dos dados multiespectrais foi conduzido no software Agisoft Metashape Professional (versão 1.6.3, Agisoft LLC, São Petersburgo, Rússia), seguindo o fluxo de trabalho recomendado para sensores MicaSense, como o RedEdge-MX. O procedimento adotado foi baseado nas orientações do manual técnico da Agisoft (2024), que descreve as etapas desde a importação das imagens até a exportação dos produtos finais, incluindo a calibração radiométrica com painel de reflectância.

Inicialmente, todas as imagens obtidas, incluindo as do painel de calibração, foram inseridas no projeto como um sistema multicâmera. O software foi configurado para

reconhecer automaticamente as imagens do painel com base nos metadados, organizando-as em diretórios específicos e aplicando máscaras automáticas para isolar apenas a área do painel. Quando a detecção automática não foi possível, a seleção e o mascaramento foram feitos manualmente. A calibração da reflectância foi realizada a partir dos valores de referência fornecidos no certificado do painel, ativando as opções *Use reflectance panels* e *Use sun sensor* para assegurar a correção radiométrica conforme as especificações do sensor RedEdge-MX (Agisoft, 2024).

Após essa etapa, as imagens foram alinhadas para gerar a nuvem de pontos esparsa e determinar as posições relativas das câmeras. Em seguida, foram inseridos os Pontos de Controle Terrestre (GCPs), cujas coordenadas foram obtidas por receptores GNSS e processadas previamente no serviço de Posicionamento por Ponto Preciso (PPP) do IBGE, com adoção do referencial geodésico SIRGAS 2000 (IBGE, 2023). Os GCPs, importados em formato CSV, foram marcados manualmente nas imagens correspondentes, garantindo o georreferenciamento preciso do modelo fotogramétrico.

O modelo foi então otimizado por meio dos parâmetros de calibração da câmera, prosseguindo-se com a geração da nuvem de pontos densa, modelo digital de elevação (MDE) e ortomosaico. Para a produção do ortomosaico, o MDE foi utilizado como superfície de referência, e o modo de mistura de imagens foi desabilitado, evitando interpolações artificiais (Agisoft, 2024). Os produtos finais foram exportados no formato GeoTIFF (.tif), apresentando resolução espacial média de 1,24 cm/pixel.

2.3. Cálculo dos Índices de Vegetação

O cálculo dos índices de vegetação foi realizado por meio de um script em Python, executado diretamente no console do QGIS (versão 3.28.5, QGIS Development Team, 2023), utilizando a API PyQGIS. O código foi desenvolvido com o objetivo de automatizar o processamento e minimizar a necessidade de intervenção manual, garantindo maior eficiência, padronização e reprodutibilidade dos resultados.

As imagens multiespectrais, exportadas individualmente por banda, foram nomeadas com prefixos padronizados (b1, b2, b3, b4 e b5), correspondendo, respectivamente, às bandas azul, verde, vermelha, red edge e infravermelho próximo (NIR). O script foi programado para percorrer automaticamente o diretório onde as imagens estavam armazenadas, identificar os arquivos com base nesses prefixos e organizar as bandas em listas para posterior combinação espectral.

Cada índice de vegetação foi calculado pixel a pixel, conforme as fórmulas implementadas no código e as bandas espectrais de entrada correspondentes. Os resultados foram exportados em formato GeoTIFF georreferenciado, contendo no nome do arquivo o título do índice e o identificador da data de coleta, o que facilitou a organização e o controle dos produtos gerados.

Esse procedimento foi repetido para cada uma das seis campanhas de voo, possibilitando a geração automatizada dos índices espectrais e eliminando a necessidade de cálculos manuais repetitivos. Os índices utilizados, NDRE (Normalized Difference Red Edge Index), TGI (Triangular Greenness Index) e VARI (Visible Atmospherically Resistant Index), foram selecionados por representarem os principais parâmetros espectrais associados ao vigor vegetativo, à senescência foliar e à cobertura do dossel na cultura da soja. As fórmulas e bandas utilizadas estão apresentadas na Tabela 2.2.

Tabela 2.2. Fórmulas e bandas utilizadas para o cálculo dos índices espectrais.

Índice de vegetação	Equação
Índice Visível Resistente à Atmosfera	$VARI = \frac{G - R}{G + R - B}$
Índice de Diferença Normalizada do Red Edge	$NDRE = \frac{NIR - RE}{NIR + RE}$
Índice de Verde Triangular	$TGI = G - 0,39 * R - 0,61 * B$
NIR (Infravermelho próximo), R (Vermelho), G (Verde), B (Azul) e RE (RedEdge) ²	

2.4. Processamento dos Dados

Os índices de vegetação utilizados neste estudo foram derivados dos ortomosaicos multiespectrais previamente processados. Para cada cultivar, calculou-se a média de reflectância com base em dez amostras pontuais de pixel, representando a assinatura espectral característica da vegetação. O experimento foi conduzido com três repetições por cultivar no campo experimental, dispostas de modo a minimizar efeitos de interferência entre genótipos adjacentes. A fim de assegurar a independência espectral das medições, consideraram-se exclusivamente os valores correspondentes à repetição central de cada parcela.

A extração dos valores espectrais foi realizada no software QGIS, utilizando-se o complemento *Point Sampling Tool*. Em cada parcela, foram selecionados dez pontos localizados no interior da área vegetada, evitando-se pixels mistos com solo exposto ou transições de borda. A média dos dez pixels foi considerada representativa da resposta espectral da cultivar, no índice, e em cada data de aquisição das imagens.

Os dados empregados nesta etapa corresponderam aos índices NDRE (Normalized Difference Red Edge), TGI (Triangular Greenness Index) e VARI (Visible Atmospherically Resistant Index), calculados a partir das imagens multiespectrais adquiridas em seis datas de voo. Inicialmente, as informações foram organizadas em planilhas independentes, correspondentes a cada data de coleta, e posteriormente integradas com base no código das parcelas, permitindo o acompanhamento da variação espectral ao longo do ciclo da cultura. As cultivares foram classificadas segundo sua finalidade (comercial e não comercial) e agrupadas em classes, precoces, médias e tardias, com base na data de floração observada em campo.

A Tabela 2.3 apresenta as variáveis mantidas após o processo de filtragem das features originais. As variáveis ID e Ciclo descrevem aspectos de identificação e classificação das cultivares, enquanto Altura, NDRE, TGI e VARI representam atributos morfológicos e espectrais da cultura. A Produtividade total foi incluída como variável agrônômica.

Tabela 2.3. Variáveis utilizadas na modelagem espectral da soja.

Variável	Descrição	Unidade	Tipo de dado
ID	Identificação única de cada cultivar	—	Categórica
Ciclo	Classificação da cultivar de soja conforme o grupo de maturação (precoce, médio ou tardio)	—	Categórica
Altura	Altura média das plantas, obtida a partir de medições de campo	cm	Contínua
NDRE	Índice de Diferença Normalizada do	—	Contínua
TGI	Índice de Verde Triangular, associado	—	Contínua

VARI	Índice Visível Resistente à Atmosfera	—	Contínua
Produtividade	Produção total de grãos por parcela experimental	g/parcela	Contínua

2.4.1. Agrupamento (Clusters)

A identificação de grupos espectralmente semelhantes foi realizada por meio do algoritmo K-Means, técnica de aprendizado não supervisionado amplamente utilizada para segmentação e classificação de dados multivariados. O número de grupos (k) foi definido a partir da aplicação do Método do Cotovelo, que avalia o comportamento da inércia, definida como a soma das distâncias quadráticas entre os pontos e seus centróides, para diferentes valores de k. A análise da curva de inércia indicou um ponto de inflexão em $k = 3$, valor adotado para a formação dos grupos, resultando nos clusters 0, 1 e 2.

A clusterização foi implementada em ambiente Python (versão 3.10), utilizando o pacote Scikit-Learn e executada no Google Colab, com suporte às bibliotecas Pandas, NumPy, Matplotlib, Seaborn e Scikit-Learn. Antes da aplicação do algoritmo, os dados espectrais foram normalizados pelo método Z-score, assegurando a comparabilidade entre variáveis com diferentes escalas de magnitude.

A clusterização foi aplicada de forma independente para cada índice espectral (VARI, TGI e NDRE). Essa abordagem permitiu identificar grupos de cultivares com respostas espectrais semelhantes ao longo do ciclo inicial da cultura e avaliar sua relação com as variáveis altura média das plantas, produtividade total e ciclo de floração.

Após a definição dos agrupamentos, foram calculadas as médias e desvios-padrão de cada variável dentro dos clusters, bem como a proporção percentual de cultivares precoces, médias e tardias em cada grupo.

A metodologia adotada segue princípios similares aos aplicados por Nounangnonhou et al. (2025), que utilizaram o algoritmo K-Means para segmentação de dados espaciais e espectrais, demonstrando sua robustez na identificação de padrões complexos e na redução de dimensionalidade em estudos agrônômicos e ambientais.

2.4.2. Análises descritivas

Para cada índice espectral e cluster identificado, calcularam-se estatísticas descritivas de tendência central e dispersão, permitindo a caracterização quantitativa dos grupos quanto à resposta espectral e ao desempenho produtivo. As variáveis altura média e produtividade média foram incorporadas às análises, possibilitando a integração entre o comportamento espectral e as variáveis agronômicas. Além disso, a proporção de cultivares precoces, médias e tardias em cada cluster foi utilizada como indicador fenológico, revelando o predomínio de determinadas fases do ciclo da soja em grupos com padrões espectrais distintos. Essa integração entre métricas espectrais e fenológicas permitiu estabelecer correspondências entre os índices de vegetação e os estádios de desenvolvimento da cultura.

3. RESULTADOS

A aplicação do algoritmo de clusterização K-Means ($k = 3$) aos índices espectrais VARI, TGI e NDRE, calculados a partir das imagens multiespectrais obtidas em seis datas (22, 43, 64, 95, 104 e 123 dias após o plantio), resultou na formação de agrupamentos distintos de cultivares de soja. As médias e desvios-padrão dos índices em cada grupo, bem como as médias de altura, produtividade e a proporção percentual dos ciclos fenológicos (precoce, médio e tardio), são apresentadas na Tabela 2.4 e 2.5.

Tabela 2.4. Valores médios e desvios padrão dos índices espectrais por cluster e dias após o plantio.

	Cluster	22 dias	43 dias	64 dias	95 dias	104 dias	123 dias
VARI	0	$0,471 \pm 0,098$	$0,860 \pm 0,327$	$0,635 \pm 0,126$	$0,691 \pm 0,135$	$0,521 \pm 0,110$	$-0,130 \pm 0,180$
	1	$0,536 \pm 0,098$	$0,462 \pm 0,524$	$0,799 \pm 0,144$	$0,915 \pm 0,129$	$0,443 \pm 0,125$	$-0,029 \pm 0,231$
	2	$0,646 \pm 0,084$	$1,102 \pm 0,355$	$0,770 \pm 0,098$	$0,570 \pm 0,111$	$0,453 \pm 0,108$	$-0,111 \pm 0,186$
TGI	0	$0,012 \pm 0,003$	$0,094 \pm 0,041$	$0,136 \pm 0,040$	$0,155 \pm 0,039$	$0,054 \pm 0,014$	$0,012 \pm 0,011$
	1	$0,014 \pm 0,002$	$0,021 \pm 0,047$	$0,139 \pm 0,038$	$0,170 \pm 0,046$	$0,063 \pm 0,017$	$0,030 \pm 0,018$
	2	$0,013 \pm 0,002$	$0,063 \pm 0,039$	$0,211 \pm 0,049$	$0,207 \pm 0,054$	$0,081 \pm 0,017$	$0,006 \pm 0,014$
NDRE	0	$-0,009 \pm 0,051$	$0,230 \pm 0,059$	$0,514 \pm 0,060$	$0,705 \pm 0,027$	$0,398 \pm 0,057$	$0,206 \pm 0,097$
	1	$0,019 \pm 0,046$	$0,361 \pm 0,052$	$0,499 \pm 0,054$	$0,681 \pm 0,028$	$0,333 \pm 0,072$	$0,238 \pm 0,099$
	2	$0,056 \pm 0,039$	$0,352 \pm 0,062$	$0,567 \pm 0,047$	$0,715 \pm 0,036$	$0,411 \pm 0,053$	$0,135 \pm 0,104$

Tabela 2.5. Médias de altura, produtividade e distribuição de ciclos das cultivares por cluster e índice espectral.

	Cluster	Altura Média (cm)	Produtividade Média (g/parcela)	Ciclo Médio	Ciclo Precoce	Ciclo Tardio
VARI	0	41,772 ± 8,704	1048,561 ± 274,457	64,90%	12,30%	22,80%
	1	39,756 ± 7,909	1153,317 ± 312,687	65,90%	22,00%	12,20%
	2	42,761 ± 8,418	1093,000 ± 286,479	78,30%	8,70%	13,00%
TGI	0	40,772 ± 8,718	1063,754 ± 324,992	64,90%	14,00%	21,10%
	1	42,306 ± 8,905	1126,653 ± 275,083	73,50%	14,30%	12,20%
	2	41,605 ± 7,370	1091,895 ± 256,637	71,10%	13,20%	15,80%
NDRE	0	41,882 ± 7,997	1026,235 ± 330,712	64,70%	9,80%	25,50%
	1	40,706 ± 8,755	1105,333 ± 281,363	68,60%	19,60%	11,80%
	2	42,048 ± 8,626	1157,667 ± 234,387	76,20%	11,90%	11,90%

3.1. Agrupamentos derivados do índice VARI

O índice VARI apresentou amplitude espectral entre os clusters, sobretudo nas fases iniciais do ciclo da cultura, 22 e 43 DAP, Figura 2.1. O Cluster 0, caracterizado por valores iniciais moderados ($0,471 \pm 0,098$ aos 22 DAP) e um pico intermediário aos 43 DAP ($0,860 \pm 0,327$), representa cultivares com desenvolvimento vegetativo equilibrado e produtividade intermediária (1048 g/parcela).

O Cluster 1 apresentou um padrão espectral atípico, caracterizado por valores reduzidos na fase inicial ($0,536 \pm 0,098$ aos 22 DAP), seguida de queda aos 43 DAP ($0,462 \pm 0,524$) e expressiva recuperação durante a fase reprodutiva ($0,915 \pm 0,129$ aos 95 DAP). Esse comportamento pode indicar genótipos com baixa densidade foliar nas fases vegetativas, mas capazes de restabelecer a atividade fotossintética e o acúmulo de assimilados durante o enchimento de grãos, refletindo mecanismos compensatórios de redistribuição e plasticidade fisiológica semelhantes aos descritos por Poudel et al. (2025), que observaram recuperação parcial da produtividade e estabilidade fisiológica em plantas de soja submetidas à desfolha moderada nas fases iniciais, devido à capacidade de compensação fotossintética e manutenção das relações fonte-dreno.

O Cluster 2 apresentou valores elevados já nas fases iniciais ($0,646 \pm 0,084$ aos 22 DAP e $1,102 \pm 0,355$ aos 43 DAP), indicando cultivares de vigor vegetativo precoce, majoritariamente pertencentes aos ciclos médio e precoce (78,3%). Entretanto, a produtividade desse grupo (1093 g/parcela) não foi a mais alta, sugerindo que o excesso de biomassa foliar inicial pode ter restringido o particionamento de fotoassimilados para

órgãos reprodutivos. Esse comportamento está de acordo com as observações de Raza et al. (2021), que verificaram que, quando a soja cresce em condições muito favoráveis, ocorre excesso de folhas no dossel. Esse aumento de folhagem provoca o sombreamento das camadas inferiores, reduzindo a entrada de luz, radiação fotossinteticamente ativa e, consequentemente, a eficiência da fotossíntese e o transporte de assimilados para as vagens e sementes. De acordo com os autores, a formação de um dossel excessivamente denso transforma as folhas em órgãos dreno, reduzindo a taxa fotossintética das folhas inferiores e antecipando sua senescência, o que compromete o rendimento final. Assim, embora o crescimento vegetativo vigoroso possa refletir maior acúmulo inicial de matéria seca, o balanço entre crescimento foliar e eficiência de partição torna-se determinante para o desempenho produtivo da soja, sendo a manutenção de um índice de área foliar ótimo essencial para maximizar o fluxo de assimilados para as estruturas reprodutivas Raza et al. (2021).

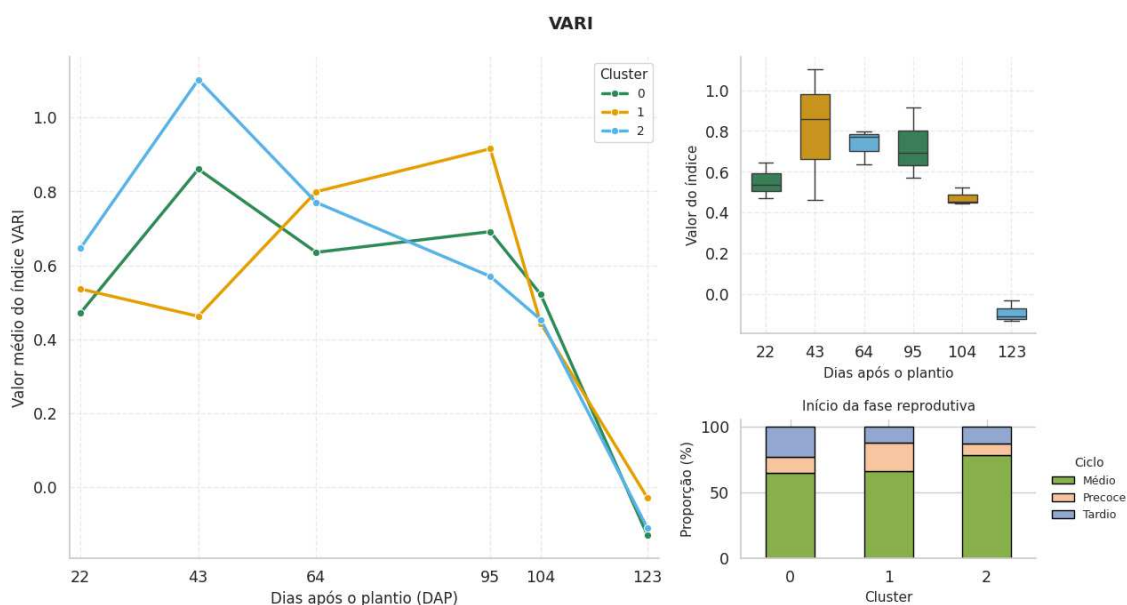


Figura 2.1 – Variação temporal do Índice VARI para os três clusters de cultivares de soja. À esquerda, as linhas representam a média ($\pm$ desvio padrão) dos valores do índice ao longo dos dias após o plantio (DAP); no canto superior direito, o boxplot mostra a distribuição dos valores por data; e no canto inferior direito, a proporção (%) de cultivares por classe de ciclo (precoce, médio e tardio) em cada cluster.

3.2. Agrupamentos derivados do índice TGI

O Índice de Verdor Triangular (TGI), Figura 2.2, apresentou valores absolutos baixos, com variações discretas entre os clusters, mas consistentes com o incremento

gradual da concentração de clorofila ao longo do ciclo fenológico da soja. O Cluster 0 evidenciou trajetória ascendente até os 95 DAP ($0,155 \pm 0,039$), seguida de leve redução na maturação, comportamento que reflete o padrão espectral típico de acúmulo e declínio de pigmentos fotossintéticos à medida que se inicia o processo de senescência. Esse comportamento é coerente com o que foi observado por Lemes et al. (2022), que destacaram a alta sensibilidade do TGI ao conteúdo de clorofila, apontando uma eficiência em detectar variações no vigor foliar e em diferenciar tecidos saudáveis daqueles em processo de estresse ou senescência, uma vez que reduções nos valores do índice estão diretamente associadas à diminuição da atividade fotossintética e da integridade foliar.

O Cluster 1 apresentou os maiores valores médios de TGI em todas as datas avaliadas, com destaque para o pico aos 95 DAP ($0,170 \pm 0,046$). Esse grupo concentrou cultivares de ciclo médio (73,5 %) e produtividade média superior (1126 g/parcela), indicando maior persistência da coloração verde e manutenção da atividade fotossintética durante o enchimento de grãos, fase determinante para o acúmulo de biomassa e formação do rendimento. De forma semelhante, Yuan et al. (2019) observaram que as variações de cor e textura do dossel obtidas por imagens RGB, especialmente por índices como o TGI, estão diretamente relacionadas ao vigor foliar e à eficiência fotossintética, permitindo antecipar diferenças de produtividade e maturidade entre genótipos de soja.

O Cluster 2 apresentou valores elevados de TGI aos 64 e 95 DAP ($0,211 \pm 0,049$ e $0,207 \pm 0,054$, respectivamente), seguidos de acentuada redução aos 123 DAP — comportamento indicativo de perda antecipada da pigmentação foliar. Esse padrão é compatível com cultivares de ciclo precoce, cujos processos de senescência ocorrem de forma mais rápida, reduzindo a duração da atividade fotossintética. Resultados semelhantes foram relatados por Miranda et al. (2025), que destacaram o potencial de índices espectrais como o TGI para identificar variações fisiológicas associadas à maturação e ao tempo de ciclo em genótipos de soja. Adicionalmente, Kipshakbayeva et al. (2024) demonstraram que a precocidade está geneticamente associada à presença de alelos recessivos nos loci E1–E4, responsáveis pela insensibilidade ao fotoperíodo e pela redução do tempo de maturação. Esses autores destacam que tais genótipos, embora apresentem menor duração do ciclo, mantêm desempenho produtivo satisfatório em ambientes de menor disponibilidade térmica, confirmando o caráter adaptativo e fisiologicamente eficiente das cultivares precoces de soja.

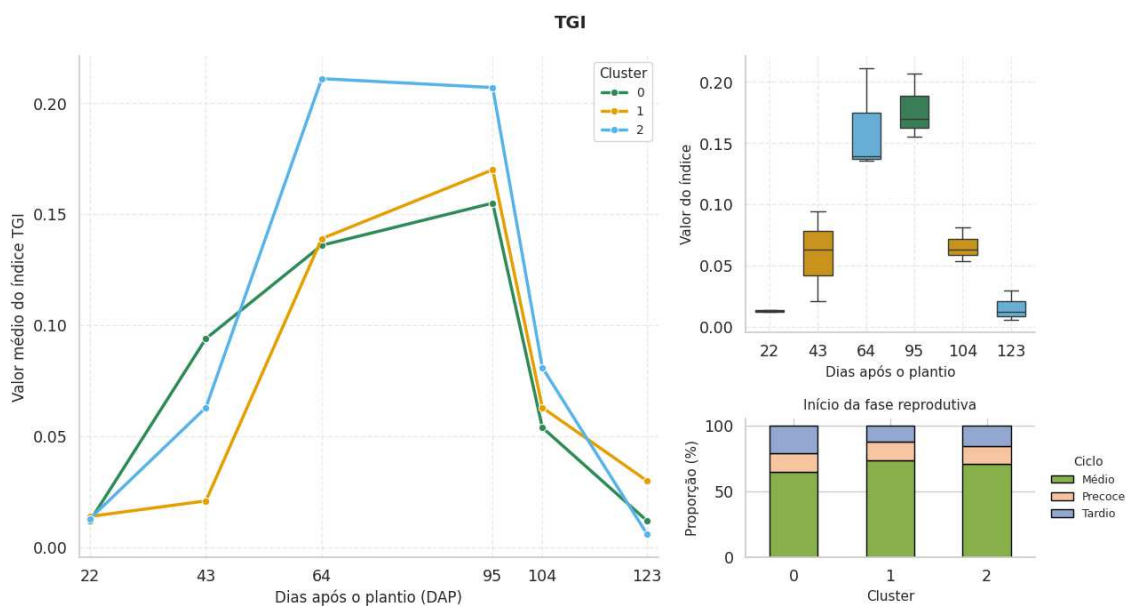


Figura 2.2 – Variação temporal do Índice TGI para os três clusters de cultivares de soja. À esquerda, as linhas representam a média ($\pm$ desvio padrão) dos valores do índice ao longo dos dias após o plantio (DAP); no canto superior direito, o boxplot mostra a distribuição dos valores por data; e no canto inferior direito, a proporção (%) de cultivares por classe de ciclo (precoce, médio e tardio) em cada cluster.

3.3. Agrupamentos derivados do índice NDRE

O NDRE, Figura 2.3, apresentou comportamento espectral estável e fisiologicamente coerente ao longo do ciclo fenológico da soja. O Cluster 0 apresentou a maior amplitude temporal do NDRE, variando de $-0,009 \pm 0,051$ aos 22 DAP para $0,705 \pm 0,027$ aos 95 DAP, seguida de redução acentuada aos 123 DAP ($0,206 \pm 0,097$). Destaque para a presença de 25,5% de genótipos tardios, e a amplitude do índice sugere que o grupo passou por um prolongado período vegetativo, o que, embora aumente a duração da fotossíntese, também tende a prolongar o ciclo e atrasar a senescência. Esse padrão é consistente com os resultados de Carneiro et al. (2020), que observaram que o NDRE apresenta menor variabilidade temporal e maior estabilidade espectral ao longo do desenvolvimento da soja, sendo menos suscetível à saturação que o NDVI em estádios avançados de crescimento. Segundo os autores, os valores de NDRE aumentam até as fases de maior vigor vegetativo (45–60 DAP), mantendo correlação significativa com biomassa, altura e teor de clorofila, antes de declinar gradualmente com o avanço da maturação.

O Cluster 1 exibiu comportamento espectral intermediário e relativamente estável ao longo do ciclo, com incremento gradual do NDRE até o máximo de $0,681 \pm 0,028$ aos

95 DAP e posterior declínio moderado na maturação ($0,238 \pm 0,099$ aos 123 DAP). Embora a estabilidade do NDRE indique uniformidade fisiológica e fotossintética durante a maior parte do ciclo, a redução antecipada dos valores do índice sugere o início precoce da senescência foliar neste grupo. De acordo com Carneiro et al. (2020), o NDRE tende a manter comportamento estável e baixa variabilidade temporal em cultivos de soja até o ponto de saturação fotossintética, sendo sensível às alterações fisiológicas que precedem a senescência, especialmente pela relação direta com o teor de clorofila e o acúmulo de biomassa. Assim, o declínio observado após 95 DAP indica redução progressiva do conteúdo de clorofila e da eficiência de absorção de radiação, condizente com o avanço para estádios reprodutivos finais

Estudos de fenotipagem multitemporal reforçam essa interpretação. Magar et al. (2025) observaram que índices baseados em red-edge, como o NDRE, respondem precocemente às alterações morfofisiológicas, refletindo a transição do balanço entre crescimento vegetativo e reprodutivo. Nessa fase, cultivares fisiologicamente equilibradas podem priorizar a alocação de assimilados para o enchimento de grãos, o que explica a senescência antecipada sem perdas acentuadas de produtividade, demonstrada pela média de 1105 g/parcela. Portanto, o padrão do Cluster 1 representa genótipos com eficiência fotossintética moderada e ciclo fisiológico mais curto, que mantêm estabilidade espectral durante o crescimento, mas antecipam o processo de senescência como estratégia adaptativa. Esse comportamento pode estar associado a cultivares de ciclo médio, capazes de otimizar o balanço entre duração do ciclo e acúmulo de biomassa, garantindo produtividade estável mesmo com menor longevidade foliar, conforme discutido por Rotundo et al. (2011), que destacam a importância do suprimento e da partição de assimilados na regulação do enchimento de grãos e na duração da atividade foliar.

O Cluster 2 destacou-se por apresentar os maiores valores de NDRE durante praticamente todo o ciclo fenológico, atingindo $0,715 \pm 0,036$ aos 95 DAP, o que reflete o elevado acúmulo de clorofila e vigor foliar entre os grupos. Esse comportamento está associado à predominância de cultivares de ciclo médio (76,2%), cuja maturação tende a ser mais homogênea e fisiologicamente eficiente. Além disso, foi o cluster com maior altura média (42,048 cm) e maior produtividade média (1157,7 g/parcela), o que sugere que a manutenção de altos valores de NDRE nas fases intermediárias do ciclo esteve diretamente relacionada ao melhor desempenho agrônômico. O subsequente declínio do NDRE na fase final ($0,135 \pm 0,104$ aos 123 DAP) representa o decréscimo natural da

clorofila e o início do processo de senescência fisiológica, em concordância com Carneiro et al. (2020), que identificaram o NDRE como o índice mais estável e sensível à variabilidade fisiológica da soja entre 45 e 60 dias após a semeadura, período de máximo acúmulo de biomassa e clorofila.

Os valores mais elevados de NDRE registrados nas fases de 64, 95 e 104 DAP mostram períodos de alta concentração de pigmentos fotossintéticos e vigor foliar sustentado, indicando a manutenção da eficiência fotossintética durante o enchimento de grãos. Esse comportamento explica o desempenho produtivo superior do Cluster 2, composto majoritariamente por cultivares de ciclo médio e fisiologia equilibrada. Segundo Miller et al. (2018), valores elevados e persistentes de NDRE ao longo do ciclo estão fortemente correlacionados ao acúmulo de biomassa e ao rendimento final da soja, uma vez que o índice expressa de forma sensível a densidade de clorofila e o vigor vegetativo do dossel. De forma complementar, Sakurai et al. (2022) demonstraram que o monitoramento multitemporal do NDRE permite identificar fases críticas de acúmulo de biomassa e prever o desempenho produtivo com alta acurácia, destacando o papel da banda red-edge na detecção da eficiência fotossintética em diferentes estádios fenológicos.

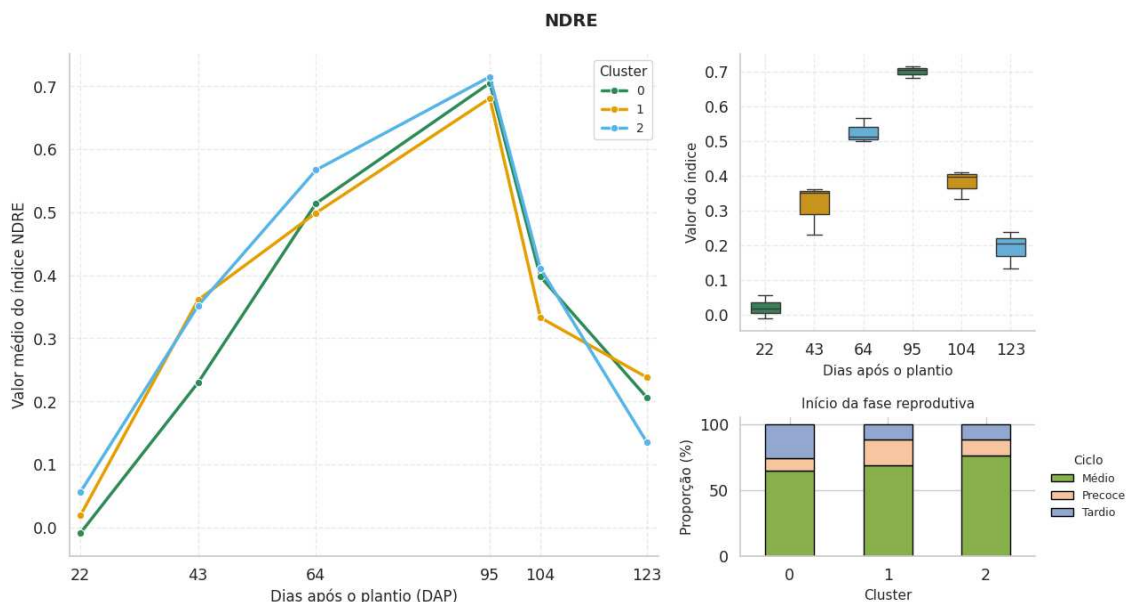


Figura 2.3– Variação temporal do Índice NDRE para os três clusters de cultivares de soja. À esquerda, as linhas representam a média ($\pm$ desvio padrão) dos valores do índice ao longo dos dias após o plantio (DAP); no canto superior direito, o boxplot mostra a distribuição dos valores por data; e no canto inferior direito, a proporção (%) de cultivares por classe de ciclo (precoce, médio e tardio) em cada cluster.

4. DISCUSSÃO

A análise multitemporal dos índices espectrais NDRE, TGI e VARI permitiu caracterizar a dinâmica espectral da cultura da soja ao longo do ciclo fenológico. A aplicação do algoritmo de agrupamento K-means auxiliou na padrões distintos de resposta espectral entre os clusters, diretamente associados às diferenças no ciclo de maturação, na eficiência fotossintética e no desempenho produtivo das cultivares. Esses resultados confirmam a capacidade dos índices derivados de sensores ópticos embarcados em VANTs de distinguir variações genotípicas e fenológicas, corroborando evidências relatadas por Guo et al. (2021) e Mendonça et al. (2022) quanto à aplicabilidade da fenotipagem remota em programas de melhoramento e manejo de precisão.

4.1. Dinâmica espectral do VARI e desempenho vegetativo inicial

O Índice Visível Resistente à Atmosfera (VARI) apresentou maior variação espectral nas fases iniciais do desenvolvimento da cultura, ressaltando sua sensibilidade à densidade foliar e à expansão do dossel. O Cluster 0, caracterizado por valores intermediários de VARI e produtividade média de 1048,6 g/parcela, refletiu cultivares com crescimento vegetativo equilibrado e eficiência moderada na conversão de energia radiante em biomassa.

O Cluster 1 apresentou comportamento espectral não linear, com decréscimo aos 43 DAP seguido de recuperação expressiva durante o enchimento de grãos. Tal padrão representa genótipos capazes de compensar deficiências vegetativas iniciais por meio do incremento da taxa fotossintética nas fases reprodutivas. Esse comportamento é consistente com o descrito por Poudel et al. (2025), que verificaram a manutenção da produtividade em plantas submetidas à desfolha parcial devido à capacidade compensatória do metabolismo fotossintético.

O Cluster 2, por sua vez, apresentou valores elevados de VARI desde os 22 DAP, caracterizando cultivares de vigor precoce e dossel denso. No entanto, a produtividade inferior (1093,0 g/parcela) indica que o crescimento vegetativo excessivo resultou em sombreamento intradossel e limitação da radiação fotossinteticamente ativa nas camadas inferiores, reduzindo a eficiência líquida da fotossíntese. Esse fenômeno foi descrito por Raza et al. (2021), que destacaram que o excesso de biomassa foliar induz à competição por luz, antecipando a senescência e comprometendo a partição de assimilados para as

estruturas reprodutivas. Dessa forma, o comportamento espectral do VARI revela que o desempenho agrônomo da soja depende do equilíbrio entre vigor vegetativo e eficiência de partição de fotoassimilados, em consonância com princípios ecofisiológicos amplamente reconhecidos. As cultivares de floração precoce apresentaram menor acúmulo de biomassa e senescência antecipada, refletida na queda precoce do NDRE e TGI após o período reprodutivo, conforme observado também por Volpato et al. (2021), que demonstraram que o declínio progressivo dos índices de verdor obtidos por UAS está diretamente associado à perda de clorofila e à definição espectral do ponto de maturidade em genótipos de ciclo curto sob condições de campo controladas.

4.2.Comportamento temporal do TGI e dinâmica de pigmentos fotossintéticos

O Índice de Verde Triangular (TGI) apresentou comportamento espectral coerente com a evolução dos pigmentos fotossintéticos durante o ciclo da soja. O Cluster 0 exibiu incremento gradual do TGI até 95 DAP ($0,155 \pm 0,039$), seguido de leve redução na maturação, refletindo a diminuição progressiva da clorofila e o avanço da senescência. Tal comportamento está de acordo com Lemes et al. (2022), que relataram que a alta sensibilidade do TGI à degradação de pigmentos e à redução da eficiência fotossintética em tecidos foliares sob estresse ou senescência.

O Cluster 1 apresentou os maiores valores médios de TGI ao longo do ciclo, destacando-se pelo pico de $0,170 \pm 0,046$ aos 95 DAP e pela produtividade superior (1126,6 g/parcela). Esse comportamento indica manutenção prolongada da coloração verde e da atividade fotossintética durante o enchimento de grãos, fase crítica para a determinação do rendimento. A correlação entre altos valores de TGI e produtividade também foi verificada por Yuan et al. (2019), que demonstraram que variações de cor e textura do dossel, derivadas de imagens RGB, são indicativas de diferenças no vigor e na maturação entre genótipos de soja.

O Cluster 2, por outro lado, apresentou redução acentuada do TGI após 104 DAP, caracterizando senescência foliar antecipada. Tal padrão é compatível com cultivares de ciclo precoce, cuja maturação fisiológica ocorre mais rapidamente. Estudos de Miranda et al. (2025) corroboram esse achado, ao demonstrar que o TGI é eficiente para detectar diferenças fenológicas associadas à duração do ciclo. De modo complementar, Volpato et al. (2021) observaram que o declínio progressivo dos índices RGB, como o TGI, está diretamente associado à maturação fisiológica e à perda de pigmentos fotossintéticos em genótipos de diferentes grupos de maturidade. Assim, o TGI se mostrou um indicador

confiável da transição entre estádios fisiológicos e da duração funcional da atividade fotossintética, em concordância também com Kipshakbayeva et al. (2024), que associaram a precocidade a alelos recessivos nos loci E1–E4.

4.3. NDRE como indicador de vigor, acúmulo de biomassa e desempenho produtivo

Dentre os índices avaliados, o NDRE (Normalized Difference Red Edge Index) apresentou o comportamento mais estável e fisiologicamente informativo. O índice mostrou-se eficaz em representar a evolução do teor de clorofila e o vigor vegetativo, sendo menos suscetível à saturação em estádios avançados, conforme descrito por Carneiro et al. (2020).

O Cluster 0 exibiu a maior amplitude temporal do NDRE, variando de $-0,009 \pm 0,051$ aos 22 DAP para $0,705 \pm 0,027$ aos 95 DAP, sugerindo prolongamento da fase vegetativa e atraso na senescência, característica típica de cultivares de ciclo tardio. Esse comportamento está associado à presença de 25,5% de genótipos tardios no grupo, corroborando a relação entre duração do ciclo e maior persistência da clorofila.

O Cluster 1 apresentou comportamento mais uniforme, com incremento gradual até $0,681 \pm 0,028$ aos 95 DAP e redução posterior na maturação ($0,238 \pm 0,099$ aos 123 DAP). A estabilidade do índice ao longo do ciclo e o declínio antecipado sugerem cultivares com balanço eficiente entre crescimento vegetativo e reprodutivo, capazes de redirecionar assimilados para o enchimento de grãos. Esse padrão é compatível com os resultados de Rotundo et al. (2011), que demonstraram que o suprimento e a partição de assimilados são fatores determinantes na regulação da senescência e na composição final das sementes de soja.

O Cluster 2 foi o grupo de melhor desempenho, apresentando os maiores valores de NDRE ($0,715 \pm 0,036$ aos 95 DAP) e maior produtividade média (1157,7 g/parcela). Esse resultado indica acúmulo superior de clorofila, manutenção prolongada da atividade fotossintética e maior eficiência assimilatória durante as fases reprodutivas intermediárias. Tais características fisiológicas foram amplamente descritas por Miller et al. (2018), que verificaram correlação positiva e significativa entre NDRE, biomassa acumulada e rendimento final, confirmando a utilidade do índice como preditor de vigor e produtividade. De forma convergente, Sakurai et al. (2022) demonstraram que o monitoramento multitemporal de índices baseados em red-edge melhora

substancialmente a acurácia na predição de biomassa e produtividade da soja, validando o uso do NDRE como indicador de desempenho fisiológico e agrônômico.

Dessa forma, os resultados obtidos neste estudo reforçam que a manutenção de valores elevados de NDRE entre 64 e 104 DAP é determinante para o aumento da produtividade, por refletir a estabilidade do aparato fotossintético e o prolongamento da assimilação de carbono nas fases críticas de enchimento de grãos.

5. CONCLUSÃO

A aplicação do algoritmo K-Means aos índices espectrais VARI, TGI e NDRE permitiu identificar grupos de cultivares de soja com respostas espectrais e fisiológicas distintas ao longo do ciclo fenológico. O VARI demonstrou maior sensibilidade às fases iniciais de crescimento e à densidade do dossel, o TGI refletiu a dinâmica e a progressão da senescência, enquanto o NDRE se destacou como o índice mais estável e informativo, apresentando forte relação com o acúmulo de biomassa, teor de clorofila e produtividade final. De modo geral, os resultados confirmam que o monitoramento multitemporal com índices espectrais obtidos por VANT é uma abordagem eficaz e precisa para avaliar a variabilidade fisiológica e produtiva da soja, permitindo distinguir grupos de cultivares com diferentes estratégias de crescimento e rendimento.

REFERÊNCIAS

AGISOFT. **MicaSense RedEdge MX processing workflow (including Reflectance Calibration) in Agisoft Metashape Professional**. Modificado em: 12 nov. 2024. Disponível em: <https://agisoft.freshdesk.com/support/solutions/articles/31000148780-micasense-rededge-mx-processing-workflow-including-reflectance-calibration-in-agisoft-metashape-pro>

ALDRIGHETTONI, J. and D'URSO, M. G.: **Advances in Precision Farming: a contribute for estimating crop health and water stress by comparing UAV Multispectral and Thermal Imagery**, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., X-G-2025.

ANZAR, S. M. et al. **Evaluation of UAV-Based RGB and Multispectral Vegetation Indices for Precision Agriculture in Palm Tree Cultivation**, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-G-2025.

CARNEIRO, F. et al. **Comparison between vegetation indices for detecting spatial and temporal variabilities in soybean crop using canopy sensors**. Precision Agriculture, v. 21, p. 1–29, 10 2020.

CHO, J. M.; LEE, B. K. **Plano de Utilização de GCP e PPK para Lidar com a Interrupção do Sinal RTK em Fotogrametria RTK-UAV**. Drones 2023.

CONAB – COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira: grãos**. v. 12, n. 11, p. 89. Brasília: CONAB, ago. 2025.

DA SILVA, M. H.; ELIAS, A. R.; DO ROSÁRIO, L. L. **Análise da cultura da soja a partir de índices de vegetação (ExG–GLI–TGI–VEG) advindos de imagens RGB obtidas com ARP**. Revista Brasileira de Geomática, v. 10, n. 2, p. 140-154, 2022.

FEHR, W. R.; CAVINESS, C. E. **Stages of soybean development**. Ames: Iowa State University, 1977. Disponível em: <https://lib.dr.iastate.edu/specialreports/87/>.

FERRAZ, M. A. J. et al. **Integrating Satellite and UAV Technologies for Maize Plant Height Estimation Using Advanced Machine Learning.** AgriEngineering 2024.

GUO, W. et al. **Uas-based plant phenotyping for research and breeding applications.** Plant Phenomics, v. 2021, p. 9840192, 2021.

HONKAVAARA, E. et al. **Digital airborne photogrammetry-a new tool for quantitative remote sensing?-a state-of-the-art review on radiometric aspects of digital photogrammetric images.** Remote Sensing, v. 1, n. 3, p. 577–605, 2009

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **IBGE-PPP: serviço on-line para pós-processamento de dados GNSS – manual do usuário.** Rio de Janeiro: IBGE, Diretoria de Geociências, Coordenação de Geodésia, jun. 2023

KIPSHAKBAYEVA, G. et al. **Assessment of the genetic parameters of soybean genotypes for precocity and productivity in the various cultivation conditions.** Heliyon, v. 10, n. 16, p. e36135, 2024.

LEMES, E. M. et al. Nascimento, F.M.A.d.; Cunha, J.P.A.R.d. **Triangular Greenness Index to Evaluate the Effects of Dicamba in Soybean.** AgriEngineering 2022.

LI, M. et al. **Crop Monitoring Using Sentinel-2 and UAV Multispectral Imagery: A Comparison Case Study in Northeastern Germany.** Remote Sens. 2022.

MAGAR, L. P. et al. **Aerial phenotyping to assess key morphophysiological traits and yield in soybean.** Smart Agricultural Technology, v. 12, p. 101276, 08 2025.

MENDONÇA, H. C. et al. **Genetic Diversity and Selection Footprints in the Genome of Brazilian Soybean Cultivars.** Front Plant Sci. 2022.

MILLER, J. J. et al. **Characterizing soybean vigor and productivity using multiple crop canopy sensor readings.** Field Crops Research, v. 216, p. 22–31, 2018.

MIRANDA, M. C. d. C. et al. **High-throughput phenotyping and machine learning techniques in soybean breeding: Exploring the potential of aerial imaging and vegetation indices.** Agronomy Journal, 2025.

NOUNANGNONHOU, C. T. et al. **GIS-based optimization of photovoltaic panel orientation in Benin using genetic algorithms and K-means clustering.** Solar Compass, Volume 16, 2025.

POUDEL, S.; MAGAR, L. P.; KHATRI, D. et al. **Identifying the most sensitive growth stages of soybean to defoliation.** Sci Rep 15, 28268, 2025.

QGIS DEVELOPMENT TEAM. *QGIS Geographic Information System: Version 3.28.5.* Open Source Geospatial Foundation Project, 2023.

RAZA, M. A. et al. **Growth Rate, Dry Matter Accumulation, and Partitioning in Soybean (*Glycine max* L.) in Response to Defoliation under High-Rainfall Conditions.** Plants 2021.

ROTUNDO, J. L.; BORRÁS, L.; WESTGATE, M. E. **Linking assimilate supply and seed developmental processes that determine soybean seed composition,**European Journal of Agronomy, Volume 35, Issue 3, 2011.

SAKURAI, K. et al. **Time-series multispectral imaging in soybean for improving biomass and genomic prediction accuracy.** The Plant Genome, v. 15, 08 2022.

SILVA, W. B. da et al. **Adaptability and stability of soybean [*Glycine max* (L.) Merrill] strains in Central-West Brazil.** Brazilian Journal of Science, v. 3, n. 7, p. 1–16, 2024.

SILVA, F. L. DA et al. **Soja: do plantio à colheita.** 2. ed. São Paulo, SP: Oficina de Textos, 2022.

VOLPATO, L. et al. **Optimization of temporal uas-based imagery analysis to estimate plant maturity date for soybean breeding.** The Plant Phenome Journal, v. 4, 05 2021.

WANG, J. A simplified processing algorithm for multi-baseline RTK positioning in urban environments. Measurement, Volume 179, 2021.

YUAN, W. et al. Early prediction of soybean traits through color and texture features of canopy rgb imagery. Scientific Reports, v. 9, 10 2019.

CAPÍTULO 3 – APRENDIZADO DE MÁQUINAS APLICADO À PREDIÇÃO DA PRODUTIVIDADE DA CULTURA DA SOJA

RESUMO

Modelos preditivos baseados em sensoriamento remoto e aprendizado de máquina apresentam elevado potencial para estimar a produtividade agrícola de forma precisa e escalável, especialmente em culturas de ampla importância econômica, como a soja. O objetivo deste estudo foi avaliar o potencial de índices espectrais obtidos por VANT e modelados pelo algoritmo Random Forest para prever a produtividade da soja e otimizar o número de coletas necessárias. Índices espectrais de vegetação (NDRE, TGI e VARI) foram extraídos de imagens multiespectrais adquiridas por VANT em seis estádios fenológicos da cultura. A base de dados incluiu 251 cultivares comerciais e não comerciais, e as estimativas de produtividade foram realizadas com o algoritmo Random Forest. Diferentes combinações temporais de coletas foram testadas para equilibrar acurácia preditiva e eficiência operacional. O modelo com todas as datas apresentou $R^2 = 0,629$, enquanto o melhor desempenho foi obtido com apenas duas coletas, realizadas aos 62 e 104 dias após o plantio (estádios R3 e R6), alcançando $R^2 = 0,7267$, MAE = 96,26 g/parcela e RMSE = 124,09 g/parcela. A combinação de três datas, 43, 95 e 123 dias após o plantio, também apresentou desempenho elevado ($R^2 = 0,7135$). Esses resultados indicam que a predição do rendimento pode ser realizada com elevada precisão mesmo com redução no número de campanhas de voo, desde que conduzidas em estádios fenológicos críticos. A abordagem proposta contribui para o avanço da fenotipagem espectral e reforça o potencial da integração entre sensoriamento remoto e aprendizado de máquina em programas de melhoramento genético e agricultura de precisão.

Palavras-chave: Soja; Sensoriamento remoto; Índices espectrais; Aprendizado de máquina

ABSTRACT

Predictive models based on remote sensing and machine learning have shown strong potential for accurately and scalable estimation of agricultural productivity, especially for crops of major economic importance such as soybean. This study aimed to evaluate the potential of spectral vegetation indices derived from unmanned aerial vehicle (UAV) multispectral imagery and modeled using the Random Forest algorithm to predict soybean yield and optimize the number of data collections required. Spectral indices (NDRE, TGI, and VARI) were extracted from multispectral images acquired by UAV at six phenological stages of soybean growth. The dataset included 251 commercial and non-commercial cultivars, and yield estimates were performed using Random Forest regression. Different temporal combinations of image acquisition dates were tested to balance predictive accuracy and operational efficiency. The model using all six dates achieved an R^2 of 0.629, while the best performance was obtained with only two image acquisitions at 62 and 104 days after planting (phenological stages R3 and R6), reaching an R^2 of 0.7267, MAE of 96.26 g per plot, and RMSE of 124.09 g per plot. The combination of three dates (43, 95, and 123 days after planting) also performed well, achieving an R^2 of 0.7135. These results demonstrate that soybean yield can be predicted with high accuracy even when reducing the number of UAV flight campaigns, provided that image acquisitions are conducted during critical phenological stages. The proposed approach contributes to the advancement of spectral phenotyping and reinforces the potential of integrating remote sensing and machine learning for data-driven decision-making in precision agriculture and breeding programs.

Keywords: Soybean; Remote sensing; Spectral indices; Machine learning

1. INTRODUÇÃO

O crescimento projetado da população mundial, que deve atingir cerca de 9,5 bilhões de pessoas até 2050 (ONU, 2019), impõe desafios sem precedentes à segurança alimentar global. Esse cenário exige a expansão da produtividade agrícola sem o aumento proporcional das áreas cultivadas, uma vez que as fronteiras agrícolas estão limitadas por fatores ambientais, climáticos e de sustentabilidade. Nesse contexto, a agricultura de precisão surge como um pilar fundamental para promover a eficiência produtiva e o uso racional de recursos naturais, apoiando-se em tecnologias de sensoriamento remoto, aprendizado de máquina e análise de dados multiespectrais.

O Brasil, maior produtor mundial de soja, consolidou-se como protagonista global na produção de grãos, alcançando uma safra recorde de 169,49 milhões de toneladas em 2025 (Conab, 2025). A importância econômica e social dessa cultura reforça a necessidade de ferramentas tecnológicas capazes de monitorar e prever a produtividade agrícola com elevada acurácia, considerando a influência combinada de variáveis como solo, clima e manejo. Nesse sentido, estudos recentes demonstram que a integração entre imagens obtidas por veículos aéreos não tripulados (VANTs) e algoritmos de aprendizado de máquina (Machine Learning) tem se mostrado particularmente promissora na estimativa de parâmetros agrônômicos e na predição da produtividade de forma rápida, não destrutiva e escalável (Yu et al., 2025).

Desde a década de 1970, o sensoriamento remoto orbital tem desempenhado papel relevante na estimativa da produtividade agrícola em diferentes escalas. Contudo, a limitação espacial e temporal de alguns satélites restringe a detecção da variabilidade intra-parcela, essencial para estudos de fenotipagem e manejo localizado. O avanço das tecnologias ópticas e multiespectrais impulsionou o uso de veículos aéreos não tripulados (VANTs), que oferecem resoluções espacial, temporal e espectral superiores às das plataformas orbitais. Essas aeronaves, caracterizadas por sua versatilidade, baixo custo e alta precisão, permitem o monitoramento detalhado de parâmetros biofísicos e fisiológicos das plantas (COLOMINA; MOLINA, 2014; MAIMAITIJIANG et al., 2017; SAGAN et al., 2019).

A combinação entre dados multiespectrais de alta resolução e técnicas avançadas de Machine Learning (ML) tem ampliado as possibilidades de monitoramento agrícola e modelagem preditiva. De acordo com Zarbakhsh et al. (2025), a aplicação de algoritmos de ML e Deep Learning (DL) tem se mostrado eficaz na identificação de padrões

complexos em dados espectrais e temporais, permitindo detectar estresses bióticos e abióticos com elevada precisão.

Nesse cenário, os modelos de aprendizado supervisionado, em especial o Random Forest (RF), têm se destacado pela capacidade de lidar com relações não lineares e conjuntos de dados de alta dimensionalidade, mantendo desempenho estável e baixo risco de sobreajuste. De acordo com Yuan et al. (2024) e Shawon et al. (2025), o RF figura entre os algoritmos mais eficazes em predições agrícolas baseadas em imagens multiespectrais de UAVs. Estudos empíricos, como o de Salami et al. (2025), reforçam que a integração entre índices espectrais fisiológicos e RF aprimora a acurácia das estimativas de produtividade de grãos, superando abordagens lineares tradicionais.

A fenotipagem aérea de alta resolução, obtida por sensores multiespectrais embarcados em drones, tem se consolidado como uma ferramenta precisa e não destrutiva para a avaliação de atributos morfofisiológicos da soja, como vigor vegetativo, senescência foliar e eficiência fotossintética (Alabi et al., 2022; Magar et al., 2025). Essa abordagem permite monitorar de forma contínua o desenvolvimento da cultura, fornecendo informações detalhadas sobre sua resposta espectral em diferentes estádios fenológicos e sob distintas condições ambientais. Ao associar medidas espectrais de alta frequência temporal e espacial a análises estatísticas e computacionais, torna-se possível compreender, com maior profundidade, os processos fisiológicos que condicionam a produtividade da cultura.

A ampla disponibilidade de combinações matemáticas entre bandas espectrais tem originado um vasto conjunto de índices de vegetação, elaborados com diferentes fundamentos teóricos e objetivos agronômicos. Muitos desses índices já demonstraram eficiência na caracterização do vigor e na estimativa da produtividade da soja (Miranda et al., 2025). Contudo, o aprimoramento dos modelos preditivos requer a seleção de variáveis com baixa correlação e maior complementaridade informacional, de modo a evitar redundâncias e capturar múltiplas dimensões do comportamento espectral da planta. Nesse contexto, os índices NDRE (Normalized Difference Red Edge Index), VARI (Visible Atmospherically Resistant Index) e TGI (Triangular Greenness Index) são relevantes por refletirem propriedades espectrais associadas a diferentes processos fisiológicos da planta. Estudos recentes destacam que a combinação de diferentes índices de vegetação, com distintos princípios de aquisição espectral, permite reduzir a redundância informacional e aprimorar a modelagem do vigor e da produtividade (Amaral et al., 2024). O NDRE (Normalized Difference Red Edge Index) tem se

destacado na cultura da soja por refletir propriedades espectrais associadas ao teor de clorofila e à eficiência fotossintética, apresentando alta correlação com o vigor vegetativo e o rendimento (Ren et al., 2023). Por sua vez, o TGI (Triangular Greenness Index) tem sido empregado na detecção de alterações pigmentares e no monitoramento do avanço da senescência, demonstrando sensibilidade a variações no conteúdo de clorofila sob condições de estresse (Lemes et al., 2022). Já o VARI (Visible Atmospherically Resistant Index) se mostra eficiente na caracterização da cobertura e densidade do dossel, sendo útil para estimar a fração de cobertura vegetal e o estágio de maturação da soja, com menor influência de fatores atmosféricos (Hu et al., 2023).

A partir dessa base espectral, o uso de modelos de aprendizado de máquina tem se mostrado particularmente promissor para a predição de produtividade agrícola. Entre os diferentes algoritmos empregados, o Random Forest (RF) destaca-se pela robustez frente à colinearidade dos dados, pela capacidade de modelar relações não lineares e pela elevada estabilidade dos resultados, mesmo em conjuntos de dados complexos e multiespectrais (Zarbakhsh et al., 2025). O RF combina múltiplas árvores de decisão produzindo estimativas médias que reduzem o sobreajuste e aumentam a precisão preditiva.

Diante desse contexto, com o presente estudo objetivou-se desenvolver um modelo preditivo da produtividade da soja com base nos índices NDRE, VARI e TGI, obtidos a partir de imagens multiespectrais adquiridas por VANT, utilizando o algoritmo Random Forest. Busca-se, assim, demonstrar o potencial dessas variáveis espectrais na estimativa não destrutiva da produtividade, explorando a complementaridade informacional entre os índices e o poder discriminante do modelo. Essa abordagem visa contribuir para o avanço da agricultura digital, fornecendo subsídios técnicos para decisões agronômicas mais precisas e sustentáveis em escala operacional.

2. MATERIAL E MÉTODOS

2.1. Área Experimental

O experimento foi realizado no Campo Experimental Professor Diogo Alves de Melo, pertencente à Universidade Federal de Viçosa (UFV), localizado em Viçosa, Minas Gerais (20°46'10.9" S, 42°52'13.8" W). A área foi destinada ao cultivo de cinco cultivares comerciais e 251 cultivares experimentais de soja, com plantio em 28 de

outubro de 2022. Para garantir o correto reconhecimento e manejo das cultivares durante todo o ciclo, foi elaborado um gabarito de plantio específico.

A área experimental foi dividida em dois blocos, sendo o Bloco 1 o foco principal deste estudo. Nesse bloco, instalaram-se 21 linhas de plantio, cada uma com 1,5 metro de largura e 0,7 metro de espaçamento entre as linhas. Cada parcela experimental foi composta por três linhas de plantio com 1,5 m de comprimento, espaçadas em 0,5 m entre si, totalizando área de 2,25 m² por parcela. As bordas do experimento foram ocupadas por cultivares de bordadura, utilizadas exclusivamente para minimizar os efeitos de borda sobre as parcelas avaliadas.

As cultivares comerciais utilizadas foram: BRASMAX Cromo TF IPRO, M5838 IPRO, NEO610 IPRO, M6410 IPRO e DM 66I68 IPRO. As cultivares não comerciais, ainda em fase experimental, foram identificadas por códigos alfanuméricos. Cada cultivar, comercial ou experimental, foi semeada por três vezes, com o intuito de minimizar a interferência entre parcelas vizinhas e aumentar a precisão das análises espectrais e agronômicas.

Todas as cultivares pertencem ao Banco de Germoplasma do Programa de Melhoramento de Soja do Departamento de Agronomia da UFV, sendo utilizadas como genitores em cruzamentos experimentais. O monitoramento da lavoura foi realizado semanalmente, com ênfase no controle fitossanitário e no manejo de pragas e doenças. As práticas de manejo seguiram as recomendações técnicas específicas para a cultura da soja, com ajustes pontuais conforme as condições observadas ao longo do ciclo (SILVA et al., 2022).

2.2. Aquisição e processamento dos dados multiespectrais

A aquisição dos dados fotogramétricos foi realizada com o VANT Matrice 100 (SZ DJI Technology Co., Ltd., Shenzhen, China), configurado para integrar sensores multiespectrais. O equipamento utilizado foi o RedEdge-MX (MicaSense Inc., Seattle, EUA), composto por cinco sensores CMOS que registram as bandas azul (475, 10 nm), verde (560, 10 nm), vermelha (668, 5 nm), RedEdge (717, 5 nm) e infravermelho próximo (840, 20 nm). As imagens foram obtidas em seis datas ao longo do ciclo da soja, entre 11h30 e 12h30, sob céu claro, para reduzir variações de reflectância (HONKAVAARA et al., 2009). Os voos foram planejados a 20 m de altitude, com 75% de sobreposição

lateral e longitudinal, e incluíram o registro do painel de calibração radiométrica para correção posterior.

Foram instalados oito Pontos de Controle em Terra (GCPs) nas extremidades do talhão (CHO et al., 2023), confeccionados em lona xadrez preto e branco ($0,30 \times 0,30$ m). Suas coordenadas foram obtidas por receptores GNSS RTK Topomap TPS T10 (Topomap, Brasil) em modo base e rover, garantindo acurácia centimétrica a milimétrica (WANG et al., 2021).

O processamento das imagens foi realizado no Agisoft Metashape Professional 1.6.3 (Agisoft LLC, São Petersburgo, Rússia), conforme o fluxo recomendado para sensores MicaSense (Agisoft, 2024). As imagens, incluindo as do painel de calibração, foram importadas como sistema multicâmera, com reconhecimento automático via metadados e máscaras aplicadas para isolar o painel. Quando necessário, a seleção foi feita manualmente. A calibração radiométrica utilizou os valores certificados do painel e as opções Use reflectance panels e Use sun sensor para corrigir as reflectâncias.

Após o alinhamento das imagens e geração da nuvem de pontos esparsa, os GCPs, processados via PPP/IBGE e referenciados ao SIRGAS 2000 (IBGE, 2023), foram inseridos manualmente, assegurando o georreferenciamento do modelo. Em seguida, procedeu-se à otimização da calibração da câmera, à criação da nuvem de pontos densa, do modelo digital de elevação (MDE) e do ortomosaico, utilizando o MDE como base e desabilitando o modo de mistura para evitar interpolações artificiais. Os ortomosaicos foram exportados no formato GeoTIFF (.tif), com resolução espacial média de 1,24 cm/pixel. As coletas foram realizadas em seis momentos distintos ao longo do ciclo da soja, abrangendo desde o estágio vegetativo até a maturação fisiológica. A Tabela 3.1 apresenta as datas de coleta e os respectivos estádios fenológicos esperados.

Tabela 3.1. Datas de coleta e estádios fenológicos da soja

Evento	Data	Dias após o plantio	Estádio esperado
Plantio	28/10/2022	0	V1
Coleta 1	19/11/2022	22	V3/V4
Coleta 2	10/12/2022	43	R1/R2
Coleta 3	29/12/2022	62	R3/R4
Coleta 4	31/01/2023	95	R5/R6
Coleta 5	09/02/2023	104	R6/R7
Coleta 6	28/02/2023	123	R8

2.3. Cálculo dos índices espectrais

O cálculo dos índices de vegetação foi automatizado por meio de um script em Python executado no console do QGIS 3.28.5 (QGIS Development Team, 2023), utilizando a API PyQGIS. O código foi desenvolvido para padronizar o processamento, reduzir intervenções manuais e assegurar eficiência e reprodutibilidade.

As imagens multiespectrais, exportadas separadamente por banda, foram nomeadas com prefixos b1 a b5, correspondentes às bandas azul, verde, vermelha, red edge e infravermelho próximo (NIR). O script percorreu automaticamente o diretório das imagens, identificou os arquivos conforme os prefixos e combinou as bandas para os cálculos espectrais.

Os índices de vegetação foram gerados pixel a pixel, conforme suas respectivas fórmulas e bandas de entrada. Cada resultado foi exportado em formato GeoTIFF georreferenciado, com o nome do índice e a data da coleta, facilitando a organização dos produtos. Esse processo foi repetido para as seis campanhas de voo, permitindo a criação automatizada dos índices sem necessidade de cálculos manuais.

Os índices espectrais utilizados, NDRE , TGI e VARI, foram escolhidos por captarem diferentes aspectos fisiológicos da soja. O NDRE está relacionado à atividade fotossintética e ao teor de clorofila, refletindo diretamente o vigor e o potencial produtivo das plantas (Ren et al., 2023). O TGI permite identificar modificações nos pigmentos foliares e acompanhar o avanço da senescência, sendo sensível às variações de clorofila em situações de estresse (Lemes et al., 2022). O VARI, por sua vez, descreve a densidade e a cobertura do dossel, auxiliando na estimativa da fração de vegetação e do estágio de desenvolvimento da cultura, com menor interferência de fatores atmosféricos (Hu et al., 2023). Na Tabela 3.2. são apresentados os índice de vegetação estudados com as respectivas equações utilizadas para calculá-los.

Tabela 3.2. Relação de IVs calculados. NIR (Infravermelho Próximo), R (Vermelho), G (Verde), B (Azul) e RE (RedEdge)

Índice de vegetação	Equação
Índice Visível Resistente à Atmosfera	$VARI = \frac{G - R}{G + R - B}$
Índice de Diferença Normalizada do Red Edge	$NDRE = \frac{NIR - RE}{NIR + RE}$

2.4. Organização, pré-processamento e avaliação das combinações temporais

Os dados espectrais e agronômicos foram inicialmente organizados em planilhas separadas para cada uma das seis datas de coleta e, posteriormente, integrados com base no código das parcelas experimentais. O conjunto de variáveis independentes (features) foi composto pelos índices espectrais NDRE, TGI e VARI, calculados a partir das imagens multiespectrais obtidas em seis momentos distintos do ciclo da soja: 22, 43, 64, 95, 104 e 123 dias após o plantio (DAP). Cada índice foi tratado separadamente em cada data, resultando em 18 variáveis preditoras (3 índices $\times$ 6 datas). A variável dependente correspondeu à produtividade total (g/parcela), medida em cada parcela experimental. A Tabela 1 apresenta a descrição das variáveis utilizadas no estudo.

Para identificar combinações temporais mais eficientes na predição da produtividade, foram testados diferentes subconjuntos de datas. Especificamente, foram avaliadas: todas as 15 combinações de duas datas entre os seis momentos de coleta; e todas as 20 combinações possíveis de três datas. Essas combinações abrangem diferentes estádios fenológicos da cultura da soja, incluindo fases vegetativas, reprodutivas iniciais, enchimento de grãos e senescência. O objetivo foi verificar quais momentos ao longo do ciclo apresentam maior potencial preditivo da produtividade final, mesmo quando utilizados em menor quantidade de coletas.

A motivação para essa abordagem reside na busca por modelos mais práticos e economicamente viáveis, especialmente no contexto da agricultura de precisão e do melhoramento genético. Para otimizar esse processo, foram avaliados subconjuntos de datas de aquisição capazes de manter desempenho preditivo comparável ao obtido com o uso de todas as coletas. Cada subconjunto foi empregado no treinamento e teste de um modelo Random Forest independente, sendo seu desempenho avaliado pelas métricas: R^2 , RMSE e MAE. Essa etapa metodológica visou identificar combinações temporais que maximizaram a relação entre desempenho e eficiência, permitindo a redução do número de voos de VANT ao longo do ciclo da cultura, a diminuição do volume de dados a serem processados e armazenados e a viabilização da aplicação prática dos modelos tanto em ambientes de produção quanto em programas de fenotipagem em larga escala.

Tabela 3.3. Variáveis utilizadas na modelagem espectral da soja.

Variável	Descrição	Unidade	Tipo de Variável
NDRE	Índice de Diferença Normalizada do	–	Contínua
TGI	Índice de Verde Triangular, associado	–	Contínua
VARI	Índice Visível Resistente à Atmosfera	–	Contínua
Produtividade total	Produção total de grãos por parcela experimental	g/parcela	Contínua

2.5. Modelagem com *Random Forest*

A predição da produtividade foi conduzida por meio do Random Forest Regressor (Breiman, 2001), implementado em Python com a biblioteca Scikit-learn (Santos et al., 2024; Alabi et al., 2022). O modelo foi ajustado com 70% dos dados para treinamento e 30% para teste, definidos aleatoriamente (`random_state = 42`) para assegurar reprodutibilidade. O RF é um método de ensemble learning baseado na combinação de múltiplas árvores de decisão, que melhora a estabilidade e reduz a variância do modelo (Zarbakhsh et al., 2025). Essa abordagem é especialmente adequada para dados multiespectrais obtidos por VANT, que apresentam alta variabilidade e correlação entre variáveis (Magar et al., 2025). Os hiperparâmetros, `n_estimators`, `max_depth` e `min_samples_leaf`, foram otimizados via validação cruzada k-fold ($k=5$) e busca em grade, garantindo equilíbrio entre desempenho e generalização (Santos et al., 2024).

2.6. Avaliação do desempenho preditivo

O desempenho dos modelos foi avaliado por meio de três métricas amplamente utilizadas na literatura: R^2 (coeficiente de determinação), expressa o quanto o modelo explica da variância da variável resposta; RMSE (*Root Mean Square Error*), mede o erro médio quadrático entre valores reais e preditos; MAE (*Mean Absolute Error*), representa o erro médio absoluto entre valores observados e estimados.

As métricas foram calculadas separadamente para cada cenário de modelagem. Os resultados foram então comparados, com o objetivo de identificar o número e o momento ideal de coletas para alcançar elevada precisão preditiva com menor custo operacional.

3. RESULTADOS

3.1. Desempenho do modelo com todas as datas

O modelo *Random Forest* treinado com os dados completos das seis datas de coleta ao longo do ciclo da soja resultou em um coeficiente de determinação (R^2) de 0,629, indicando que aproximadamente 62,9% da variabilidade na produtividade foi explicada pelas variáveis espectrais e agrônômicas utilizadas. Os valores de erro médio absoluto (MAE) e raiz do erro quadrático médio (RMSE) foram, respectivamente, de 105,82 g/parcela e 144,52 g/parcela. Na Figura 3.1 é apresentado o gráfico de dispersão entre os valores observados e preditos pelo modelo, com os pontos coloridos de acordo com a altura das plantas no florescimento.

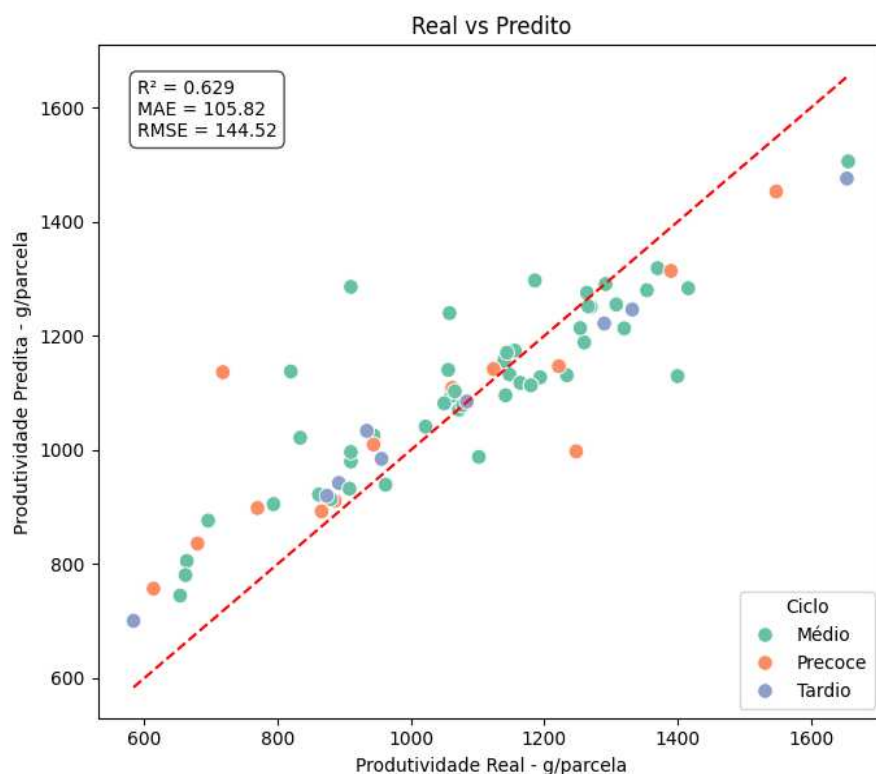


Figura 3.1. Gráfico de dispersão: produtividade real vs. predita para o modelo com 6 datas.

3.2. Desempenho dos modelos com combinações de três datas

Foram avaliadas todas as 20 combinações (Tabela 3.4) possíveis de três datas entre as seis coletas realizadas. O objetivo foi identificar subconjuntos temporais com elevada capacidade preditiva, reduzindo a necessidade de monitoramento contínuo ao longo do ciclo da cultura.

Tabela 3.4. Desempenho do modelo preditivo construído a partir de combinações de três datas de aquisição de imagens, expressos pelos coeficientes de determinação (R^2) e pelos erros médio quadrático (RMSE) e absoluto médio (MAE). Os valores de RMSE e MAE estão apresentados em gramas por parcela (g/parcela).

Coleta 1	Coleta 2	Coleta 3	R^2	MAE	RMSE
43 DAP	95 DAP	123 DAP	0,7192	98,301	125,7341
43 DAP	62 DAP	123 DAP	0,7086	99,8852	125,6363
43 DAP	104 DAP	123 DAP	0,6778	111,0881	132,1108
62 DAP	104 DAP	123 DAP	0,6708	106,3748	133,5462
95 DAP	104 DAP	123 DAP	0,6660	111,7910	134,5191
43 DAP	95 DAP	104 DAP	0,6424	111,7171	139,1829
22 DAP	43 DAP	62 DAP	0,6380	107,3981	140,0351
22 DAP	43 DAP	123 DAP	0,6324	113,2510	141,1226
43 DAP	62 DAP	104 DAP	0,6298	107,5143	141,6025
62 DAP	95 DAP	123 DAP	0,6206	113,6595	143,3550
43 DAP	62 DAP	95 DAP	0,6069	113,4490	145,9173
22 DAP	43 DAP	104 DAP	0,5987	111,2919	147,4444
22 DAP	95 DAP	104 DAP	0,5695	122,8567	152,7130
62 DAP	95 DAP	104 DAP	0,5343	119,2590	158,8272
22 DAP	43 DAP	95 DAP	0,5302	129,2090	159,5248
22 DAP	62 DAP	104 DAP	0,5280	126,2790	159,8959
22 DAP	62 DAP	123 DAP	0,5145	127,4662	162,1640
22 DAP	95 DAP	123 DAP	0,5111	127,9990	162,7358
22 DAP	104 DAP	123 DAP	0,4857	132,9433	166,9102
22 DAP	62 DAP	95 DAP	0,4227	130,8910	176,8383

A melhor combinação foi composta pelas datas de 43, 95 e 123 dias após o plantio, que correspondem, respectivamente, aos estádios fenológicos entre R1 (início da floração), R5.3 (fase intermediária do enchimento de grãos) e R8 (maturação fisiológica). Este modelo apresentou R^2 de 0,719 MAE de 98,30 g/parcela e RMSE de 125,73 g/parcela, desempenho muito próximo ao modelo com todas as seis datas (Figura 3.2). As informações das fases reprodutivas são altamente informativas para a predição de rendimento, especialmente o intervalo que compreende o florescimento e o enchimento dos grãos, períodos nos quais há forte relação entre vigor fisiológico, interceptação de radiação e produtividade final. A redução no número de campanhas de voo para três, com perda mínima de acurácia, demonstra a viabilidade de estratégias de monitoramento mais enxutas em contextos operacionais reais.

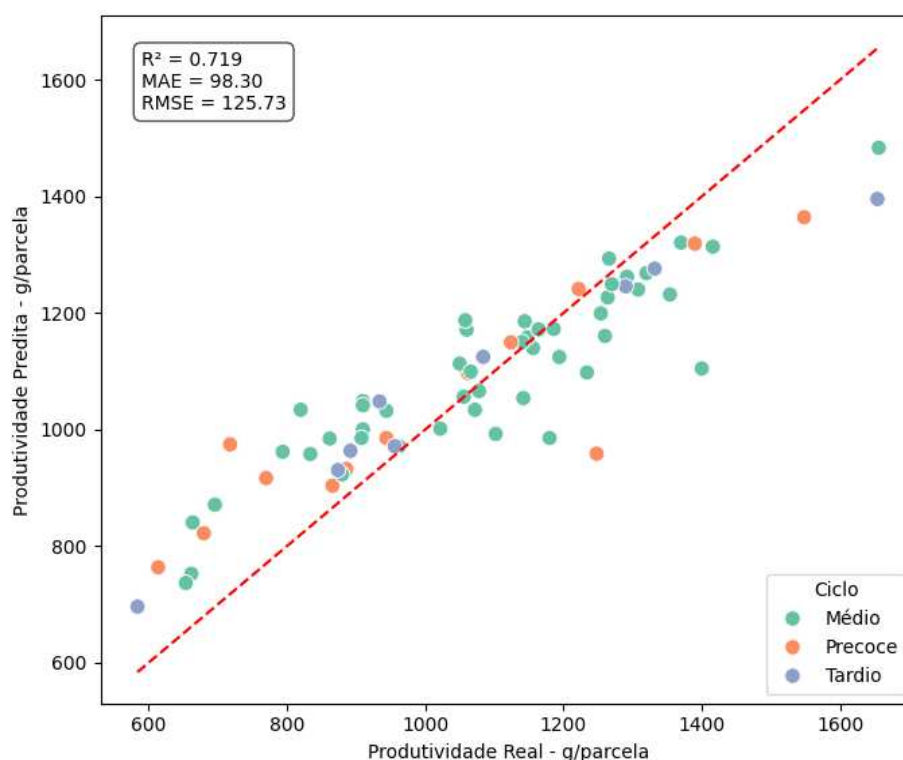


Figura 3.2. Gráfico de dispersão: modelo com melhor trio de datas. Os valores de RMSE e MAE estão apresentados em gramas por parcela (g/parcela).

3.3. Desempenho dos modelos com combinações de duas datas

Na sequência, foram avaliadas todas as 15 combinações possíveis de duas datas, Tabela 3.5, com foco na identificação de modelos ainda mais enxutos e operacionais. O

resultado mais promissor foi obtido com a combinação de 62 e 104 dias após o plantio, que correspondem entre os estádios de R3 (formação de vagens) e R6 (enchimento completo dos grãos). Surpreendentemente, esse modelo superou os demais em desempenho, com R^2 de 0,7267, MAE de 96,26 g/parcela e RMSE de 124,09 g/parcela (Figura 3.3). Esses resultados são relevantes, pois indicam que é possível alcançar acurácia preditiva superior ao modelo com todas as datas, utilizando apenas duas coletas realizadas em momentos fisiológicos críticos para a definição do rendimento da soja.

Tabela 3.5. Desempenho do modelo preditivo construído a partir de combinações de duas datas de aquisição de imagens, expressos pelos coeficientes de determinação (R^2) e pelos erros médio quadrático (RMSE) e absoluto médio (MAE). Os valores de RMSE e MAE estão apresentados em gramas por parcela (g/parcela).

Coleta 1	Coleta 2	R^2	MAE	RMSE
62 DAP	104 DAP	0,7267	96,2699	124,0956
62 DAP	123 DAP	0,6945	98,9447	131,1839
43 DAP	62 DAP	0,6837	107,6321	133,4864
43 DAP	104 DAP	0,6513	101,6005	140,1705
104 DAP	123 DAP	0,6477	108,6388	140,8773
22 DAP	62 DAP	0,6448	105,4109	141,4691
62 DAP	95 DAP	0,6429	113,1541	141,8405
95 DAP	123 DAP	0,6225	113,0948	145,8310
43 DAP	95 DAP	0,6097	116,1568	148,2973
43 DAP	123 DAP	0,6084	108,5247	148,5293
95 DAP	104 DAP	0,6082	111,9817	148,5669
22 DAP	104 DAP	0,5425	116,1709	160,5538
22 DAP	43 DAP	0,5420	119,0672	160,6405
22 DAP	123 DAP	0,5365	116,9928	161,5949
22 DAP	95 DAP	0,4942	123,4346	168,8085

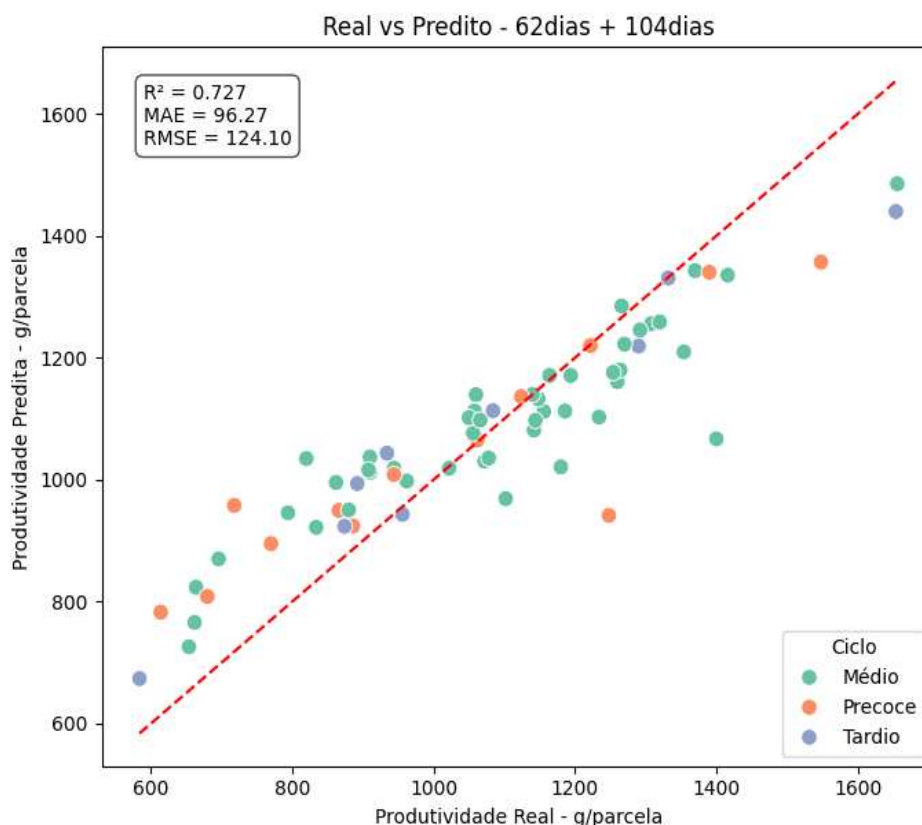


Figura 3.3. Gráfico de dispersão: modelo com melhor dupla de datas. Os valores de RMSE e MAE estão apresentados em gramas por parcela (g/parcela).

3.4. Comparação entre cenários

Na Tabela 3.6 é apresentado um resumo do desempenho dos três principais modelos: com todas as datas, com o melhor trio e com a melhor dupla. Observa-se que tanto o modelo com três coletas (43, 95 e 123 dias após o plantio) quanto o modelo com duas coletas (62 e 104 dias após o plantio) apresentaram desempenho comparável, ou até superior, ao modelo completo. O modelo baseado em apenas duas datas obteve o melhor desempenho entre todos os cenários, com o maior valor de R^2 e os menores erros (MAE e RMSE), indicando o alto poder preditivo das coletas realizadas entre os estádios R3 (formação de vagens) e R6 (enchimento completo dos grãos). Esses resultados reforçam a viabilidade de estratégias de monitoramento espectral mais enxutas, sem comprometer a acurácia da predição da produtividade.

Tabela 3.6. Comparativo entre os modelos com diferentes números de datas de coleta. Os valores de RMSE e MAE estão apresentados em gramas por parcela (g/parcela).

	Datas utilizadas	R²	MAE	RMSE
6 datas	22, 43, 62, 95, 104, 123	0,629	105,82	144,52
Melhor trio	43, 95, 123	0,719	98,301	125,73
Melhor dupla	62, 104	0,7267	96,26	124,09

4. DISCUSSÃO

Os resultados obtidos demonstram o alto potencial do uso de índices espectrais derivados de imagens multiespectrais, combinados com algoritmos de aprendizado de máquina, para a predição da produtividade da cultura da soja. A utilização do modelo Random Forest, com variáveis obtidas em diferentes estádios fenológicos, permitiu capturar com eficácia a variabilidade agrônômica entre as parcelas experimentais, mesmo em um conjunto de dados composto por cultivares de ampla diversidade genética.

O modelo com todas as seis datas de coleta apresentou bom desempenho ($R^2 = 0,629$), confirmando que a densidade temporal de dados espectrais ao longo do ciclo pode explicar uma parte substancial da variação no rendimento. Esse resultado converge com achados de Alabi et al. (2022), que observaram destaque do Random Forest entre modelos de predição multiespectral em soja.

Contudo, com os resultados foram identificados modelos mais enxutos com desempenho superior ao modelo completo. A combinação de apenas duas datas (62 e 104 dias após o plantio), correspondentes aos estágios R3 (formação de vagens) e R6 (enchimento completo dos grãos), resultou em um modelo com $R^2 = 0,7267$, MAE de 96,26 g/parcela e RMSE de 124,09 g/parcela, o melhor desempenho entre todos os cenários avaliados. Esse resultado é consistente com estudos que identificam os estádios reprodutivos R3 a R6 como fases decisivas para a predição do rendimento da soja, nas quais os modelos espectrais apresentam maior acurácia, conforme observado por Shammi et al. (2024), que destacaram o estágio R3 como o mais promissor para predição precoce e os estádios R5–R6 para estimativas tardias de produtividade.

Além da melhor dupla, a análise das 20 combinações de três datas identificou que o trio formado por 43, 95 e 123 DAP, correspondentes entre R1 a R8, também obteve excelente desempenho ($R^2 = 0,719$), indicando que a inclusão da maturação fisiológica (R8) contribui com informações adicionais relacionadas ao sucesso final da translocação de assimilados. Conforme descrito por Vogel et al. (2021), a manutenção da atividade

fotossintética e da duração da área foliar durante o enchimento dos grãos é determinante para o acúmulo de biomassa e a formação do rendimento na soja. Essa persistência fisiológica justifica a relevância espectral das coletas realizadas em estádios finais, como R8, quando ocorre o máximo acúmulo de biomassa e o dossel expressa o sinal de maturidade (Pérez et al., 2024; Miranda et al., 2025) o que justifica a relevância espectral da coleta em R8 para a previsão do rendimento final.

O NDRE, por utilizar a banda Red-edge, mostrou-se altamente informativo na identificação de pequenas variações no vigor das plantas durante os estágios reprodutivos. O TGI, por sua vez, apresentou utilidade como marcador de senescência e degradação foliar nas fases finais, enquanto o VARI se mostrou mais eficaz na detecção do fechamento do dossel nos estádios iniciais e intermediários. A complementaridade entre esses índices explica o bom desempenho dos modelos que utilizaram múltiplas datas, e principalmente, a eficácia de combinações específicas.

Do ponto de vista prático, a substituição do modelo completo (com seis coletas) por um modelo baseado em apenas duas coletas estratégicas representa um avanço considerável para a agricultura de precisão. Reduzem-se custos operacionais, processamento de imagens e armazenamento de dados, sem comprometer a acurácia da estimativa. Essa abordagem ganha especial relevância em programas de fenotipagem em larga escala, onde o custo e a logística de múltiplas campanhas de voo são limitantes. De forma semelhante ao observado por Pérez et al. (2024) demonstraram que a redução planejada no número de coletas espectrais não compromete a precisão da predição, mantendo elevados valores de R^2 e baixos erros mesmo com apenas três a cinco voos. Essa otimização operacional favorece a aplicação de métodos baseados em aprendizado de máquina em programas de melhoramento genético e agricultura de precisão. Além disso, a abordagem aqui adotada valida o uso de Random Forest como um algoritmo robusto e eficaz para modelagem da produtividade agrícola.

5. CONCLUSÃO

Os resultados confirmam o potencial do uso combinado de índices espectrais e algoritmos de aprendizado de máquina para estimar a produtividade da soja de forma precisa e eficiente. O modelo Random Forest apresentou desempenho consistente mesmo diante da alta variabilidade genética entre as cultivares avaliadas.

Destaca-se que a melhor performance foi alcançada com a utilização de apenas duas datas estratégicas de coleta (62 e 104 dias após o plantio), correspondentes aos estádios fenológicos R3 e R6. Essa combinação resultou em um coeficiente de determinação (R^2) de 0,7267, superando até mesmo o modelo baseado em todas as datas do ciclo.

Essa economia de coletas, sem perda significativa de acurácia, demonstra que é possível reduzir custos operacionais e otimizar o uso de recursos, mantendo a robustez das estimativas. O modelo proposto mostra-se promissor para aplicações em agricultura de precisão e programas de melhoramento genético, oferecendo suporte confiável à tomada de decisões agronômicas baseadas em dados.

REFERÊNCIAS

- AGISOFT. MicaSense RedEdge MX processing workflow (including Reflectance Calibration) in Agisoft Metashape Professional. Modificado em: 12 nov. 2024. Disponível em: <https://agisoft.freshdesk.com/support/solutions/articles/31000148780-micasense-rededge-mx-processing-workflow-including-reflectance-calibration-in-agisoft-metashape-pro>. 2024.
- ALABI, T. R.; ABEBE, A. T.; CHIGEZA, G.; FOWOBAJE, K. R. Estimation of soybean grain yield from multispectral high-resolution UAV data with machine learning models in West Africa. *Remote Sensing Applications: Society and Environment*, v.27, p.100782, 2022.
- AMARAL, L.; OLDONI, H.; BAPTISTA, G.; FERREIRA, G.; DE FREITAS, R.; MARTINS, C.; CUNHA, I.; SANTOS, A. Remote sensing imagery to predict soybean yield: a case study of vegetation indices contribution. *Precision Agriculture*, v.25, p.2375-2393, 2024.
- BREIMAN, L. Random Forests. *Machine Learning*, v.45, p.5–32, 2001.
- CHO, J. M.; LEE, B. K. **Plano de Utilização de GCP e PPK para Lidar com a Interrupção do Sinal RTK em Fotogrametria RTK-UAV**. Drones 2023.
- COLOMINA, I.; MOLINA, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, v.92, p.79-97, 2014.
- CONAB – COMPANHIA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira: grãos, safra 2024/25 – 9º levantamento. Brasília, DF: CONAB, 2025.
- HONKAVAARA, E.; ARBIOL, R.; MARKELIN, L.; MARTINEZ, L.; CRAMER, M.; BOVET, S.; CHANDELIER, L.; ILVES, R.; KLONUS, S.; MARSHAL, P.; SCHLÄPFER, D.; TABOR, M.; THOM, C.; VEJE, N. Digital airborne photogrammetry - a new tool for quantitative remote sensing? A state-of-the-art review on radiometric aspects of digital photogrammetric images. *Remote Sensing*, v.1, n.3, p.577–605, 2009.

HU, J.; YUE, J.; XU, X.; HAN, S.; SUN, T.; LIU, Y.; FENG, H.; QIAO, H. UAV-Based Remote Sensing for Soybean FVC, LCC, and Maturity Monitoring. *Agriculture*, v.13, p.692, 2023.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. IBGE-PPP: serviço on-line para pós-processamento de dados GNSS – manual do usuário. Rio de Janeiro: IBGE, Diretoria de Geociências, Coordenação de Geodésia, jun. 2023.

LEMES, E. M.; COELHO, L.; ANDRADE, S. L. D.; OLIVEIRA, A. D. S.; MARQUES, M. G.; NASCIMENTO, F. M. A. D.; CUNHA, J. P. A. R. D. Triangular Greenness Index to Evaluate the Effects of Dicamba in Soybean. *AgriEngineering*, v.4, p.758-769, 2022.

MAGAR, L. P.; SANDIFER, J.; KHATRI, D.; POUDEL, S.; CHILUWAL, A. UAV-based aerial phenotyping to assess key morphophysiological traits and yield in soybean. *Smart Agricultural Technology*, v.12, p.101276, 2025.

MAIMAITIJANG, M.; GHULAM, A.; SIDIKE, P.; HARTLING, S.; MAIMAITIYIMING, M.; PETERSON, K.; SHAVERS, E.; FISHMAN, J.; PETERSON, J.; KADAM, S.; BURKEN, J.; FRITSCHI, F. Unmanned Aerial System (UAS)-based phenotyping of soybean using multi-sensor data fusion and extreme learning machine. *ISPRS Journal of Photogrammetry and Remote Sensing*, v.134, p.43-58, 2017.

MIRANDA, M. C. C.; AONO, A. H.; FAGUNDES, T. G.; ARDUINI, G. M.; PINHEIRO, J. B. High-throughput phenotyping and machine learning techniques in soybean breeding: Exploring the potential of aerial imaging and vegetation indices. *Agronomy Journal*, v.117, e70012, 2025.

NAÇÕES UNIDAS NO BRASIL. População mundial deve chegar a 9,7 bilhões de pessoas em 2050, diz relatório da ONU. Brasília, DF: ONU Brasil, 2019.

PÉREZ, O.; DIERS, B.; MARTIN, N. Maturity Prediction in Soybean Breeding Using Aerial Images and the Random Forest Machine Learning Algorithm. *Remote Sens.*, v.16, p.4343, 2024.

QGIS DEVELOPMENT TEAM. *QGIS Geographic Information System: Version 3.28.5.* Open Source Geospatial Foundation Project, 2023. Disponível em: <https://qgis.org>. Acesso em 15 set. 2025.

REN, P.; LI, H.; HAN, S.; CHEN, R.; YANG, G.; YANG, H.; FENG, H.; ZHAO, C. Estimation of Soybean Yield by Combining Maturity Group Information and Unmanned Aerial Vehicle Multi-Sensor Data Using Machine Learning. *Remote Sens.*, v.15, p.4286, 2023.

SAGAN, V.; MAIMAITIJIANG, M.; SIDIKE, P.; EBLIMIT, K.; PETERSON, K. T.; HARTLING, S.; ESPOSITO, F.; KHANAL, K.; NEWCOMB, M.; PAULI, D. UAV-Based High Resolution Thermal Imaging for Vegetation Monitoring and Plant Phenotyping Using ICI 8640 P, FLIR Vue Pro R 640, and thermoMap Cameras. *Remote Sensing*, v.11, p.330, 2019.

SALAMI, Z.; BABAMURATOV, B.; AYYAPPAN, V.; PRABAKARAN, N.; KHUDAYBERGANOV, K.; SINGH, C. Modelling Crop Yield Prediction with Random Forest and Remote Sensing Data. *Natural and Engineering Sciences*, v.10, p.67-78, 2025.

SANTOS, L. B.; GENTRY, D.; TRYFOROS, A.; FULTZ, L.; BEASLEY, J.; GENTIMIS, T. Soybean yield prediction using machine learning algorithms under a cover crop management system. *Smart Agricultural Technology*, v.8, 100442, 2024.

SAROWAR, M. S.; EMA, F. B.; MAHI, A. K.; NIHA, F. L.; ZUBAIR, H. T. Crop yield prediction using machine learning: An extensive and systematic literature review. *Smart Agricultural Technology*, v.10, 100718, 2025.

SHAMMI, S.A.; HUANG, Y.; FENG, G.; TEWOLDE, H.; ZHANG, X.; JENKINS, J.; SHANKLE, M. Application of UAV Multispectral Imaging to Monitor Soybean Growth with Yield Prediction through Machine Learning. *Agronomy* 2024.

SHAWON, S. M.; EMA, F. B.; MAHI, A. K.; NIHA, F. L.; ZUBAIR, H. T. Crop yield prediction using machine learning: an extensive and systematic literature review. *Smart Agricultural Technology*, v. 10, p. 100718, 2025.

SILVA, F. L. DA et al. Soja: do plantio à colheita. 2. ed. São Paulo, SP: Oficina de Textos, 2022.

VOGEL, J. T.; LIU, W.; OLHOFT, P.; CRAFTS-BRANDNER, S. J.; PENNYCOOKE, J. C.; CHRISTIANSEN, N. Soybean Yield Formation Physiology - A Foundation for Precision Breeding Based Improvement. *Front Plant Sci.*, v.12, 719706, 2021.

WANG, J.; XU, T.; NIE, W.; XU, G. A simplified processing algorithm for multi-baseline RTK positioning in urban environments. *Measurement*, v.179, p.109446, 2021.

YU, Z.; CHEN, Q.; SHAN, Q.; LV, Y.; CHEN, Z.; ZHOU, Y.; ZHANG, P.; JIANG, H.; CAO, W. Developing soybean yield prediction model based on multicolor space and texture features of UAV images. *Computers and Electronics in Agriculture*, v.235, 110345, 2025.

YUAN, J.; ZHANG, Y.; ZHENG, Z.; YAO, W.; WANG, W.; GUO, L. Grain Crop Yield Prediction Using Machine Learning Based on UAV Remote Sensing: A Systematic Literature Review. *Drones*, v.8, p.559, 2024.

ZARBAKSH, S.; FAKHRZAD, F.; RAJKOVIC, D.; NIEDBAŁA, G.; PIEKUTOWSKA, M. Approaches and challenges in machine learning for monitoring agricultural products and predicting plant physiological responses to biotic and abiotic stresses. *Current Plant Biology*, v.43, 100535, 2025.

CONCLUSÃO GERAL

Os resultados obtidos ao longo desse trabalho confirmam a viabilidade técnica e científica da utilização de índices espectrais derivados de sensores multiespectrais embarcados em VANTs, associados a algoritmos de aprendizado de máquina, para a predição da produtividade da soja. As análises realizadas demonstraram que a correlação entre índices varia ao longo do ciclo da cultura, sendo possível identificar combinações com baixa redundância informacional e maior complementaridade espectral.

A evolução temporal dos índices NDRE, TGI e VARI revelou padrões distintos entre grupos de maturação e níveis de produtividade, destacando-se o potencial desses índices na caracterização fenotípica de cultivares. A modelagem com Random Forest indicou que coletas realizadas nos estádios reprodutivos críticos, especialmente entre R3 (formação de vagens) e R6 (enchimento de grãos), proporcionam capacidade preditiva com redução do número de observações, otimizando recursos operacionais.

A melhor combinação identificada utilizou apenas duas datas (62 e 104 dias após o plantio), alcançando R^2 de 0,7267, superando inclusive o modelo com seis coletas ao longo do ciclo. Os resultados auxiliam na confirmação da hipótese de que é possível reduzir o número de campanhas de monitoramento sem comprometer a acurácia, o que favorece a adoção prática da metodologia em programas de fenotipagem em larga escala e na agricultura de precisão.

Conclui-se, portanto, que a integração entre sensoriamento remoto multiespectral e aprendizado de máquina constitui uma abordagem promissora para o avanço da agricultura digital, permitindo a tomada de decisão mais precisa e eficiência no manejo e no melhoramento genético da soja.